IN-74-CR
104050

JPL Publication 90-2

P.64

Developments in Fiber Optics for Distribution Automation

H. Kirkham
H. Friend
S. Jackson
A. Johnston

December 1, 1991

Prepared for

Office of Energy Storage and Distribution
United States Department of Energy

Through an agreement with

National Aeronautics and
Space Administration

by

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

JPL Publication 90-2

Developments in Fiber Optics for Distribution Automation

H. Kirkham
H. Friend
S. Jackson
A. Johnston

December 1, 1991

Prepared for

Office of Energy Storage and Distribution
United States Department of Energy

Through an agreement with

National Aeronautics and
Space Administration

by

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Prepared by the Jet Propulsion Laboratory, California Institute of Technology, for the U.S. Department of Energy through an agreement with the National Aeronautics and Space Administration.

This publication reports on work performed under NASA Task RE-152, Amendment 203 and sponsored through DOE/NASA Interagency Agreement No. DE-AI01-79ET 29372 (Mod. A009).

ABSTRACT

An optical fiber based communications system of unusual design is described. The system consists of a network of optical fibers overlaid on the distribution system. It is configured as a large number of interconnected rings, with some spurs. Protocols for access to and control of the network are described. Because of the way they function, the protocols are collectively called AbNET, in commemoration of the microbiologists' abbreviation Ab for antibody. Optical data links that could be optically powered are described. There are two versions: each has good frequency response and minimal filtering requirements. In one, a conventional FM pulse train is used at the transmitter, and a novel form of phase-locked loop is used as demodulator. In the other, the FM transmitter is replaced with a pulse generator arranged so that the period between pulses represents the modulating signal. Transmitter and receiver designs, including temperature compensation methods, are presented. Experimental results are given.

ACKNOWLEDGMENTS

The work described in this report is part of a continuing effort funded by the U.S. Department of Energy. We would particularly like to recognize the contribution of Mr. K.W. Klein, Director of the Office of Energy Storage and Distribution of DOE. As this report was being finalized, Ken announced his retirement from DOE. Ken has had an interest in distribution automation and the application of fiber optic technology to power systems for a long time. His support at OESD began the whole effort, and his continuing involvement and guidance has helped us maintain momentum.

PREFACE

This report is one of a series describing work performed at the Jet Propulsion Laboratory for the United States Department of Energy on the technology of distribution automation.

In the preface to our previous report, written about a year ago, I said that it was the judgment of the United States Department of Energy and other industry observers that in the mid-1990s the United States would experience a generation capacity shortage. Opinions have not changed. There is talk of non-utility generation, and of wheeling power, but there is no conviction that these alone will suffice. Load management, and perhaps some of the equipment utilization functions of distribution automation, can help ameliorate the effects of a capacity shortage, and seem certain to be required within a few years. Consequently, both DOE and JPL consider that continued involvement in this distribution automation effort is contributory to the solution of a problem of national importance.

Distribution automation continues to be the subject of research, at JPL and elsewhere, and as yet there are no large-scale commercial applications. There are several reasons for the lack of commercial success, and these are themselves the subject of much discussion. Distribution automation is a complex topic. Consider three aspects: time-scale, control strategy and communications requirements.

The various distribution automation functions have time-scales ranging from fractions of a second (for protection) through minutes (most control functions) to hours (thermal time-constants) or even years (load growth planning). Some of the functions require centralized control or decision-making (load reduction); others can be implemented in a distributed fashion (VAr control). Sometimes a centralized function may need to override a local one; load reduction and voltage control are examples. Some functions require communications to many locations in order to be effective, others can take advantage of local intelligence. It is likely that the very complexity of distribution automation has weighed against its widespread implementation. There are problems to be solved in the area of power system control, communications networks, optimization, and customer relations, as well as safety and economics, often considered overriding factors.

Broadly speaking, the examples of distribution automation that do exist have been implemented piecemeal, as the need for a function arose, as technologies advanced or as funds became available. Rarely do such assemblages of subsystems constitute a system, operating with coordination. There are problems, both technical and institutional, of considerable complexity. We have adopted the approach of separating the technical (and economic) problems from the institutional ones, and then analyzing the technical problems and their economic constraints.

Because it has until now operated in a communications-limited environment, the distribution automation community has developed some ingenious approaches, and the line between control and communications problems has become blurred. Our approach and our analysis have segregated the fundamental functional requirements of distribution automation. We perceive distribution automation as a mixture of control (for operations) and planning functions. Data acquisition operates in support of these. Communications is not a "requirement": but some functions require centralized monitoring, and these depend on communications.

Poor communications has been one of the factors responsible for the slow growth of the distribution automation market. Power line carrier, radio and telephone--the "conventional" communications media--have been inadequate. Over the last year or so we have been working at applying fiber optics to solve the communications problem at minimum cost. But a fiber optic communications system for distribution automation must be more than just economical: it must reach all parts of the distribution system, and do so reliably. A network of unusual design is proposed as the solution. Some possible solutions to the special problems of this network are discussed in this report.

The overall communications network, called AbNET in this report, has far-reaching implications for the topic of distribution automation. While the operating protocols for the network are distributed, the applications programs are all assumed to reside at the distribution substation. Distribution automation is assumed to be "centralized," at least to this extent. To be effective, interfaces at the substation and at the remote terminals will have to be defined carefully. But in return, the strategy offers an end to vertical integration in the industry. Once this communications system has been installed, the customer need no longer depend on the supplier of the communications system for the requisite control software. There is no reason that software to operate the distribution system cannot be bought from one of a number of competing vendors, or developed by the utility itself. Furthermore, such changes to such software can be implemented simply by installing them at the distribution substations. A significant impediment to progress in distribution automation is removed.

A further impediment to the widespread application of the techniques of distribution automation has been the high cost of data, particularly the high cost of measurements. The distribution system is electrically very complex, with multiple loads, taps, capacitors and switching arrangements. For full-scale distribution automation, complete instrumentation of the system may be required, calling for literally hundreds of measurements on a single feeder. Even assuming that the communications system required for all this information is in place, the sensing needs are an economic burden to the system.

In earlier work, we demonstrated an optically powered data link that could make a contribution to reducing the cost of measurements in the distribution system. In this report, we present two new versions of the link, that offer improved performance in terms of requiring less filtering, while maintaining good frequency response.

In the future, JPL may be involved in the development of suitable hardware and software for the implementation of one or more prototype distribution automation systems using fiber optics. Some of that work may be industry-funded: one of the purposes of this report is to introduce potential sponsors to the JPL effort.

HAROLD KIRKHAM

Pasadena, California
November, 1989

TABLE OF CONTENTS

LIST OF FIGURES

Figure | Page

LIST OF FIGURES (Continued)

1. FIBER-OPTICS BASED COMMUNICATIONS FOR DISTRIBUTION AUTOMATION

1.1. Introduction

There is growing interest in distribution automation, the implementation of automation in the electric power distribution system. Some of the communications problems associated with distribution automation are discussed in an earlier report (Kirkham, Johnston and Friend, 1989). This section of this report describes recent developments aimed at solving those problems. The primary problem, limited data rate on the communications channel, is readily solved by using fiber optics. However, as shown by Kirkham *et al. (ibid.)*, a rather unusual communications system must be used.

It must be assumed that it is necessary to communicate with all parts of the power network. Consequently, the communications network must be co-extensive with the power system. The data acquisition and control needs of the system mean that it will be necessary to tap into the communications network at a large number of locations. (For even a single substation, this number could be hundreds.) It is also a requirement that communications be possible to all locations even if the configuration of the network changes, for example because of failure of one or more lines or nodes.

Normally, the distribution system is *operated* radially, with a limited set of open loops designed to provide an alternative way of bringing power to any given location. The fiber can, of course, cross an open power switch. As a result, the fiber optic communication system is arranged not as a conventional ring, star or bus system, but as a series of interconnected loops, with an occasional spur.

An example of a system for power distribution communications is shown below in Figure 1-1. In the Figure, a *node* is a remote terminal unit (RTU) that could be involved in data acquisition or control, or both. A *gateway* is a special kind of node that normally prevents messages that originate in one substation from reaching the area served by another station, but that can be programmed to pass such messages, if required. The lines connecting the nodes are assumed to be full-duplex communications lines.

The highly interconnected nature of the communications system means that most locations could be accessed from more than one direction, and over several different routes. These are features that will help provide reliable communication, but which complicate network access. This is a somewhat unusual topology for a fiber-optic local-area network, in that there are a large number of interconnected loops with common legs, and many branch points.

The large number of points at which the fiber must be accessed for data acquisition or control means that ordinary optical taps, in which the power in the fiber is divided evenly, cannot be used. What is more, even if the sensors used for data acquisition are optical, there is practically no chance of the signals from them being compatible with the information format used in the communications fiber. Therefore, each tap point will necessarily include optical-electrical-optical conversion.

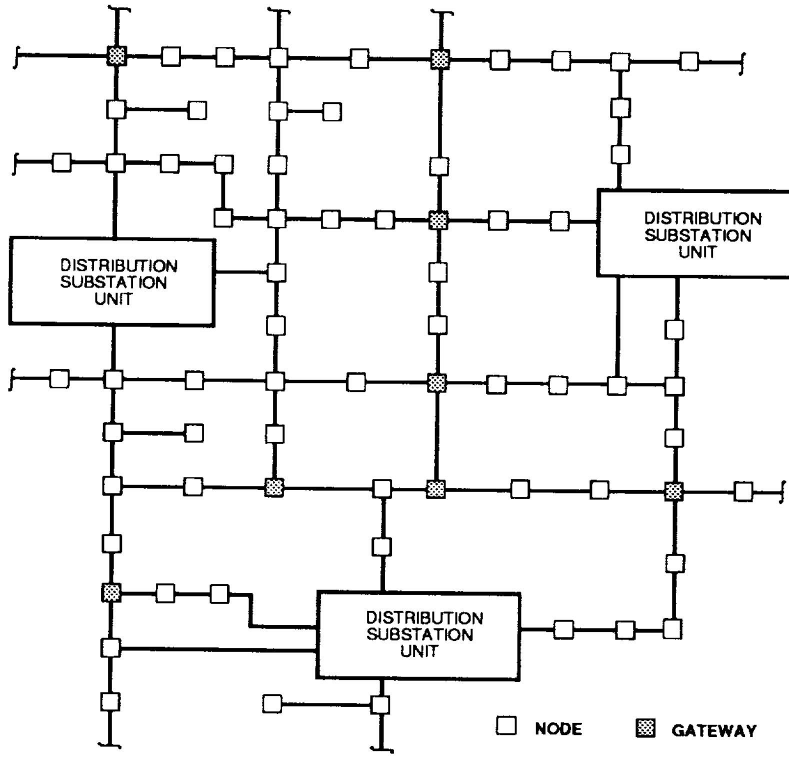


Figure 1-1. Topology of communications system for power distribution network

Figure 1-2 is a block diagram of the node required to implement the communication system. The use of a microprocessor to control the flow of data in and out of the communications system is assumed, although it is possible that simple functions such as load management could be performed by a purely hardware system. For completeness, a general-purpose node that is capable of data acquisition and control signal generation is shown.

Altogether, a rather unusual system is defined. As far as communications are concerned, the system is rather liberally sprinkled with repeaters. The network consists of a large number of point-to-point links, with the stations effectively connected in series. This means that communication distances are very short, and even the slight dispersion effects that might appear because of the use of low-grade fiber will not accumulate in the system.

There is no shared medium or common bus. Consequently, most messages will have to pass through many intermediate stations on their way to their destination. The operation of such a communications system can be controlled in a number of ways. In the next section, we will review the question of distribution automation communications, and place the problem in the wider context of computer communications.

1.2. Review of Communications for Distribution Automation

There are a number of media that have been used for communication in the distribution system. None was ideally suited to all applications. Distribution line carrier (DLC) can be used to implement load management, or some such low-speed function. The data rate (typically a few bits per second) is so low that messages must usually be broadcast, there being insufficient capacity to include addressing. Telephone can be used, but leased lines are point-to-point, and many lines would be required to reach all the places that distribution automation must access. The public telephone system can be used to reach consumers' premises or to communicate with distribution substations. In either application, the cost of service from the telephone company is likely to make the application difficult to justify economically. There are communications methods that use radio, but two-way communication requires two sets of equipment. In any case, there is a very limited amount of spectrum space available.

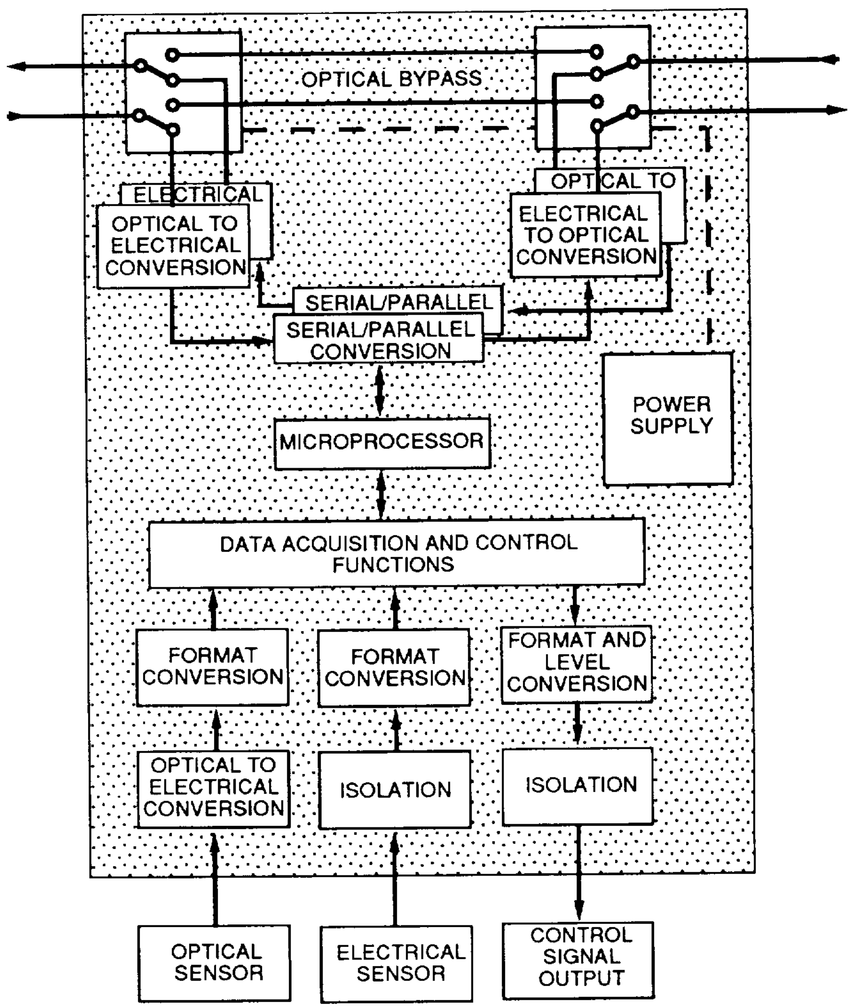


Figure 1-2. Block diagram of node

These technical and economic limitations, together with institutional issues such as ownership and the need for licenses, have been impediments to the implementation of distribution automation. Each time a utility wanted to add a distribution automation function, they had to consider a new communications system for it. They might, for example, have successfully automated the control and monitoring of a distribution feeder by means of a radio system, and then been required to add DLC to perform load management on the same feeder. Probably the radio system did not provide access to all the loads, or it may have been too expensive to modify the control software.

The result of this approach has been that electricity distribution communications has avoided a problem that faced most other users of communications systems. Most communications systems are able to support more than one user. In order to share the limited resource (the communications channel), the users have to abide by some agreed-to protocol. By and large, there is no need for such protocols in utility communications. Each user or application has its own dedicated communications system.

In contrast, computer communications networks have been obliged to address the problem. Practically any physical medium can be used to allow just two computers to communicate. One may choose twisted pair, coaxial cable or fiber optics, depending on the speed requirements. The two users can agree on a data encoding method for the medium (baseband or carrier, signal levels etc.) and communications can begin. As soon as a third or fourth computer is added, additional questions have to be answered. If it is decided not to connect all machines to all other machines, and not to connect all machines to a common point, there are questions of routing and access to the common medium that cannot be left unanswered.

A series connection means that some messages must pass through an intermediate machine to reach their destination, but avoids the problem of contention for access to the common channel of a parallel connection. If a series connection is used, should the computers be arranged in a ring? Is a bus interconnection the right way to go? If the system uses a common communications medium, as in a bus, how is the access contention problem resolved? The addition of more users to the system brings more decisions that have to be made. A hierarchical structure, such as a tree network, might be useful. In a highly interconnected network, how are messages to be routed?

As computer networks developed, the need for standards for interconnection became apparent. In this way, the location of other users and the manufacturer of their equipment could be made transparent to a user. The Open System Interconnection (OSI) model of the International Standards Organization (ISO) (Zimmermann, 1980) was a recent step towards standardization. This model split the problem of computer communications into seven logical and physical layers. While other arrangements are possible, and indeed were under study when the ISO model was released, the OSI model has become the accepted framework for discussing computer communications.

1.3. Open System Interconnection Model

Computer-to-computer communications in the ISO seven layer model is shown in Figure 1-3. In the Figure, a few words of explanation have been added to the diagram. A more detailed explanation is given below. It is important to realize that this OSI standard is just a model. Very few networks adhere strictly to this structure. Usually layers are missing because they are not needed in some particular application, and sometimes functions are implemented in layers other than those to which this model would ascribe them.

To make the model seem more familiar, we use the problem of getting two personal computers to talk to one another as an example. We begin this description with the top layer and proceed downward. Imagine that you have a personal computer with, say, a word-processing program. You want to send a file from your word processor to that of a colleague.

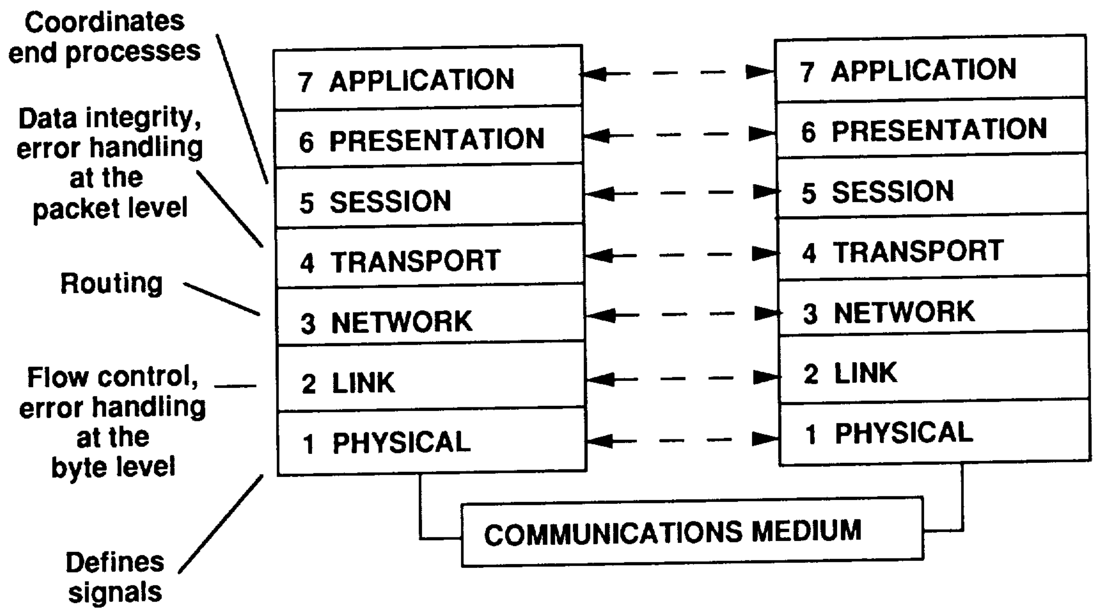


Figure 1-3. International Standards Organization model of open system interconnection

The application layer is the layer of software that transfers data across the network. All other layers in the hierarchy exist solely to satisfy the needs of this layer. In the discussion example, this is the word processor program. In the case of a distribution automation system, the applications layer might be the SCADA program, the program that presents data to the operator, or the program that contains the control algorithms which are commonly part of SCADA operations.

The presentation layer provides services to the application layer to process the data in some way to make them more suitable for the layers below. This could mean, for example, translation or encryption of the data. The operating system of the personal computer is an example of a presentation layer. It furnishes data in a format that can be handled, for example, by the word processing program or the screen driver, and at the same time can take a file of data furnished by a word processing program and present it for transmission to a remote computer to the layer underneath, the session layer. In the distribution automation example, this layer could contain user-callable library routines.

The session layer is the first of the layers in this hierarchy specifically concerned with communications to another computer. In essence, this layer is responsible for coordinating interaction between the opposite end application processes. The layer has to be aware of both the application and the communications. Some of the functions of the typical communications program on a personal computer perform the functions required of the session layer. For example, a program sending blocks of data to make up a file might pause for acknowledgement at the end of each block. This could be a session layer function. In some applications, the session layer is a "virtual" layer. The decisions made here are the type of communication to be employed (e.g. full or half duplex) and how the failure of lower layers in the hierarchy is to be handled.

The session layer interfaces to the <u>transport layer</u> which is the highest of the seven layers responsible for the integrity of data. For example, this layer would be capable of performing error checking, perhaps on a packet by packet basis. This function is performed in the personal computer world by a number of communication programs.

The job of the transport layer is to furnish error-free messages, in sequence, to the session layer. Whether this is a simple task or a complex one depends on the layers beneath. If the lower layers are retaining message sequence, and are performing error checking, the transport layer becomes very simple. On the other hand, if the lower layers can operate so as to get messages out of sequence, for example, the transport layer must correct the deficiency.

The transport layer connects to the <u>network</u> layer, which is responsible for furnishing data to the bare-bit manipulations of the data link layer below. In the PC world, the network layer is typically part of the telephone system, not the PC itself. The network layer is responsible for routing the information from computer to computer. Once you dial the phone number, you have no choice over the route that the phone company uses for your data. Consequently, it is their responsibility in this particular case to perform the functions of the network layer.

The <u>data link</u> layer is the highest level at which information as such is handled. The data link layer may perform error control on a word by word basis; for example, parity checking occurs in the data link layer. There may also be some means of flow control or hand-shaking to assure synchronization between devices capable of operating at different speeds. The telephone company for example, in their data link layer, typically will add bits to the 8-bit word coming from most PC modems, so that error checking (in addition to the parity bit that the user knows about) is usually performed.

Some users of the model have divided the data link layer into two sub-layers: <u>logical link control</u> and a <u>medium access control</u>. The logical link sub-layer is responsible for establishing, maintaining and terminating a logical connection between devices. The medium access control sub-layer ensures that only one device attempts to transmit at a time, performing the function of congestion control.

The <u>physical layer</u>, the lowest layer of this hierarchy, is the level at which the electrical signals are exchanged. A specification of the physical layer typically includes a description of electrical and mechanical quantities involved. For example, RS-232 describes the type of plug to be used as well as the voltage levels and the pin connections. In the world of modems, the physical layer includes a description of the procedures to establish or release connections between electrical circuits such as phone lines.

For most purposes in exchanging data from one computer to another, it is only the bottom three layers of this hierarchy that need to be standardized. However, because these layers do need to be standardized for effective communication between different machines, a good deal of work has gone on in this area. IEEE, for example, has developed a series of standards according to the IEEE 802 Committee. These are usually known by the decimal organization of the subcommittees, for example, IEEE 802.3 is CSMA/CD (Carrier Sense Multiple Access with Collision Detection) and IEEE 802.5 is Token Ring.

One may note also that the functional relationships between layers are clearly defined in the open system interconnection model, but are not so clearly defined in practice. The visibility of the boundaries between layers need only be clearly defined if the boundary corresponds to a product, in which case interfacing to that product (either physically or in terms of software) will be vastly simplified. Thus, the transport layer and the session layer functions are both performed adequately by PC communications programs; and the network layer, data link layer, and physical layer occur outside the boundaries of the typical modem.

1.4. Communications Issues for Distribution Automation

In arriving at the mesh configuration shown in Figure 1-1, we made the assumption that a single communications system would be used for all possible distribution automation applications. The system would therefore have to reach all parts of the distribution system, and be capable of handling all future control and monitoring data. The distribution system topology defined the communications network topology, and the data requirements led us to choose fiber optics as the communications medium. This, in turn, leads to the choice of a baseband system, probably using differential Manchester encoding. The need for multiple taps led to an arrangement of node-repeaters, rather than passive optical power splitters. While the system size (length of links, number of nodes) and topology are thus defined, there are some remaining system issues.

In a conventional communications system design, link data rate is determined by traffic requirements. For distribution automation, the data rate required is very low compared to the capability of the channel, and we can instead let cost considerations fix the data rate. While a final determination has not yet been made, we expect a bit rate of about 10 Mb/s to be achievable at minimum cost. Since the medium is fiber optics, the expected error rate can also be fixed by design. (Error rate is influenced by power margin and bit rate, but is unaffected by external factors.) As is common practice in fiber optics communications systems, a bit error rate (BER) of 10^{-9} is assumed.

This still leaves unanswered a number of questions of system operation. In terms of the seven-layer model, we have so far defined only the lowest. We have not addressed the questions of how to avoid collisions, whether to use error checking to ensure reliable communications, how to route messages in the network, or whether voice and data can be transmitted over the same system. There are also questions of priority (should some users be assigned a higher priority than others?), and the general question of whether to adopt a distributed strategy for operation or a centralized one.

Generally speaking, these questions are addressed by considering the system application. A telephone trunk application, for example, would indicate a design with high efficiency, in terms of the fraction of the time that the communications channel was carrying (revenue earning) data. Efficiency, in this sense, is affected by the channel bit rate (Mb/s), the packet length (assuming a packet-switching network) and the length of the trunk. A real-time application might be more concerned with <u>access delay</u>. Perhaps because it was not explicitly considered at the early stages of design, the access delay in some experimental distribution automation systems routinely amounts to tens of seconds, and is sometimes measured in minutes!

While transmission delay is likely to be important in many distribution automation functions, in our application the most important factor is likely to be <u>reliability</u>. One of the goals of distribution automation is to improve the reliability of service by improving the performance of the distribution system. It must be presumed that a reliable communications system is a prerequisite. Ideally, the communications system should be failure-resistant, and preferably quite immune to single-contingency failures. The usual strategy for ensuring reliable communications at the lower levels of the hierarchy is error detection and request for retransmission. At the higher levels re-routing can be used. For the time being, assume that an effective error detection scheme is used to mitigate the rare occurrences of corrupted messages. What should happen if the network is changed by the operation of a bypass at a node, or by failure of a link? In other words, how should the network layer be used for maximum reliability?

If the substation unit maintains a map of the distribution system configuration, messages can be routed by a minimal number of intermediate nodes to any desired location. This would minimize transmission delay. Routing information for the response could even be included in the message. However, if the map is inaccurate (perhaps because of a line or node failure), this method breaks down, and some other strategy must be used. It may be possible for the central unit rapidly to update the map, but this approach seems cumbersome. We feel that any method that relies on a "retry" is inappropriate. Better to flood all paths in the communications network with information, than to use a routing approach. It is possible that this approach may lead to lower efficiency, but it is our conviction that performance will still be within acceptable bounds for accomplishing distribution automation.

We may consider the communications network for distribution automation to consist of a number of interconnected rings in which information can be circulated. A conventional communications system with the stations connected in series would consist of only one ring. For this kind of network, communications protocols based on passing a token (which is a series of bits conferring the right to transmit) from station to station have been worked out. An example is ANSI/IEEE Standard 802.5-1985. The kind of network that can be controlled by IEEE 802.5 protocol is shown below in Figure 1-4.

In this kind of system, a token is circulated in the ring, passing from unit to unit until one of them wishes to transmit information. When this occurs, that unit removes the token from the ring and replaces it with its message. At the end of this process, the token is placed back on the ring and circulated behind the message, so that other units can access the ring if they want to. In order to prevent the endless circulation of the same information, the message is removed by an "active monitor" in the ring. This is a station that is specifically checking a particular bit of information (known as the monitor bit) in the system overhead data to solve this problem. The active monitor can remove a message from the ring, and thereby prevent its endless circulation.

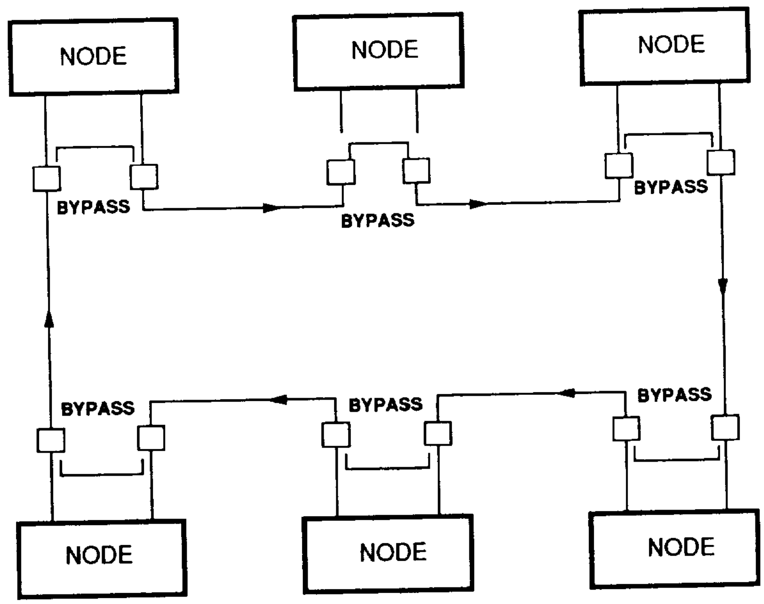


Figure 1-4. Token Ring system for ANSI/IEEE 802.5

1.5. Adapting a Token-Passing Ring Protocol: The Problems

The token passing strategy does not readily adapt to the multiply-interconnected rings or mesh configuration required for access to all portions of the communications network for power distribution. At least as of 1986 there were no commercially available local networks based on the interconnected mesh (Hopper, Temple and Williamson, 1986, page 40). There are two problems:

First, the token can be duplicated in a system that includes branches. This can occur because, in order to maximize communications reliability, a node at an intersection of two rings is expected to insert any information it receives into both rings. Both rings might thus contain a copy of a valid token. Because of this, a message could be generated simultaneously by a node in each ring. At a point where two such rings come together, a collision (in the sense of simultaneous messages from two directions) would occur. Such a node must then include some means of buffering one message while another one is being processed. This is not a difficult problem to solve.

The second problem is that a single active monitor cannot prevent message circulation in a multiple ring system. Further, if the topology of the system is capable of changing during operation, it is impossible *a priori* to place one and only one active monitor in each ring. To solve this problem, each node must be able to terminate the process of repeating a message. A new, distributed protocol is required in order to solve this problem.

1.6. Distributed Network Control

Distributed strategies for solving the message-circulation problem have been discussed in the literature, and some have been implemented in practice. It seems, however, that no such strategy has been described for the multiple ring problem. In fact, there has been little work on multiple-ring networks, and what work has been done has assumed that the configuration is based on geographically small areas served by rings, interconnected and communicating only occasionally. Our system is rather different because its basic configuration is that of the power system. The multiple rings associated with one distribution substation are expected to communicate routinely to maximize communications reliability: occasional communication with an adjacent substation may also be necessary.

1.7. The Solution: AbNET

The problems of adapting a token ring system to mesh topology are mainly problems of the network layer of the hierarchy. Therefore, we describe the solution at this level first.

1.7.1. AbNET Network Layer

In order to effect a distributed strategy to solve the circulation problem, every message must contain a unique identifying number that can be stored in every node that it passes through. Since the operation of nodes in separate rings is assumed to be quite independent, the information required to identify a message uniquely could be based on the source address of the originating node. This address could be a "logical" address, or it could be simply the street address of the node. The identifier probably need only contain one or two additional bits, because it is unlikely that a greater number of separate messages could be simultaneously generated in any given ring. Alternatively, the unique identifier could be simply the source address and the time of origination of the message. The unique identifier numbers can readily be stored in a small "stack" in the node's memory. We will show later that the higher layers of the hierarchy provide an even simpler way of identifying messages, but for now imagine that each message is somehow uniquely identified.

At all nodes, any message received is retransmitted on all outgoing lines, unless the message has been seen before, or is addressed to the node in question. By adopting such a strategy, a message inserted anywhere into the communications network shown in Figure 1-1 will be broadcast to all units in the network, and will not be repeated by a given node more than once. All links transmit the message exactly once. The method is therefore as economical in operation as a single ring with an active monitor. It is in some ways more efficient than a method which relies on some central unit working out a route for messages to go from node to node. The method, known as flooding, always chooses the shortest path between two nodes, because it chooses all paths. Transmission delay is therefore minimized.

Reliability is also maximized. The scheme will work even if a fiber is broken, and a node is therefore disconnected from one of its neighbors. (Such a failure can be detected by communications tests that could be ordered from time to time from the substation unit, but this is not an essential feature of normal operation.)

The communications protocol resembles the working of the body's immune system. The first time a T-cell in the immune system is exposed to an invading organism, it learns to recognize the organism as "non-self." On a second exposure, the T-cell will produce antibodies that kill the invader. In our communications network, the messages are the invading organism, and they wander throughout the network as far as available communications channels will allow. On first exposure to a message, our nodes store information that will allow them, on a second exposure, to kill the message. Because of this similarity, one of us (Friend) has proposed we call the system "AbNET," after the microbiologists' abbreviation "Ab," for antibody.

The flowchart of Figure 1-5 describes in broad terms how the network layer functions of an AbNET node could be implemented in software. The procedure that starts at the top of the flowchart assumes the existence of a valid (error-free) message in an input buffer. The data link layer functions that accomplish this are discussed next.

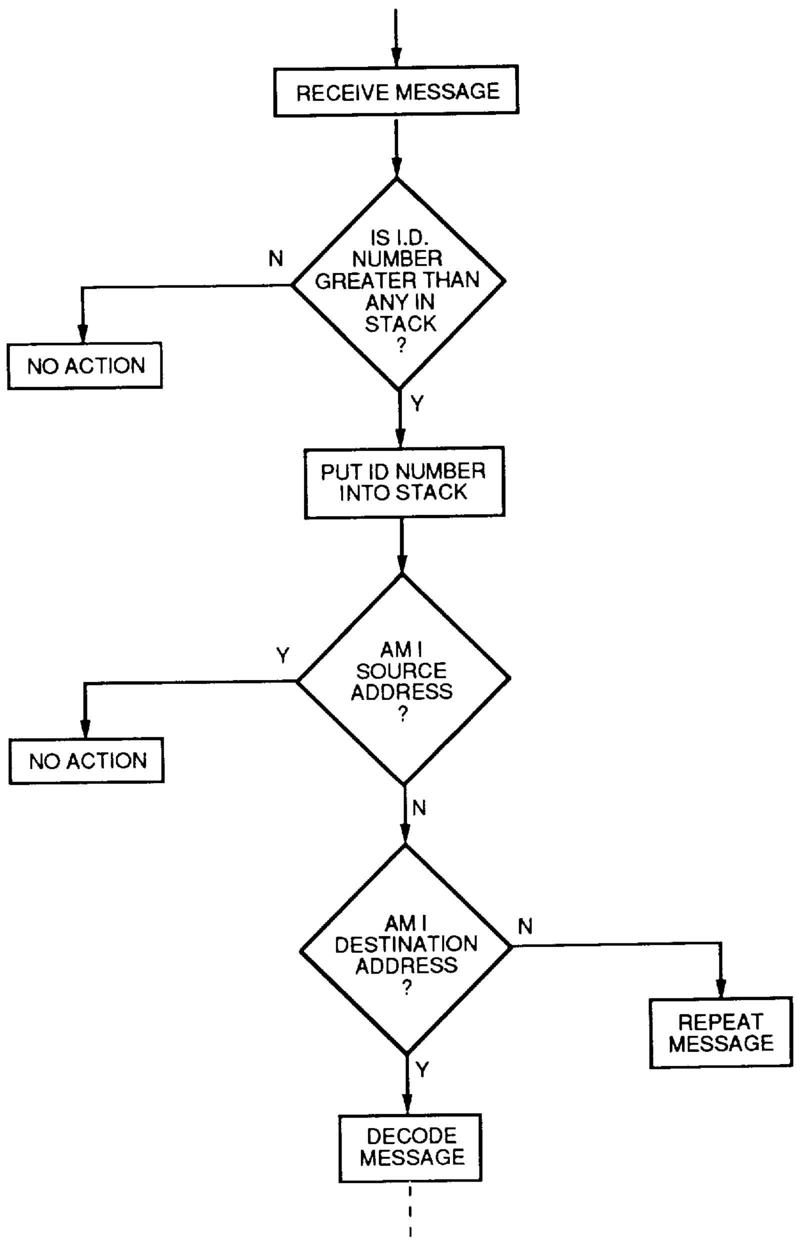


Figure 1-5. AbNET network layer flowchart

1.7.2. AbNET Data Link Layer

Normally, the data link layer is responsible for delimiting data fields, acknowledgement of receipt of data, and some error control, such as parity check. In most communications systems, receipt of information that passes the error check is acknowledged to the sending station. The detection of an error will normally cause a request for retransmission. In addition to this kind of function, the data link layer may contain a flow control mechanism, to prevent problems when two devices of different speeds try to communicate. (One can hear this process when two modems exchange tones to establish a data rate.)

Some simplifications are possible in AbNET. First, the bit rate of all nodes can be fixed in advance, so flow control should not be needed. A data encoding scheme such as differential Manchester can be used, so that receiving devices can synchronize themselves to the incoming data, and so that there is no net dc level in the signal (an important consideration with optical communications).

There seems to be little to be gained by retaining the conventional technique of message receipt acknowledgement. With a BER of 10^{-9}, the probability of sending a message M bits long without error is $P = (1-10^{-9})^M$. If the message is 1000 bits long, the success probability is 99.99990%. In other words, only one message in a million would require retransmission or correction. It seems inherently wasteful to devote any amount of effort to acknowledging successful reception of data. We propose instead the use of an error correcting code that would reduce the occurrence of uncorrected errors to insignificance.

While the details of an AbNET specification have not been completed, and will be addressed by our group in the next year, it seems likely at this point that uncorrectable communication problems in a fiber optics network may indicate a more severe problem than could be addressed by either the data link or network layers. This topic is examined later. For now, note that the absence of a retransmission-request procedure means that all traffic in the network is associated with the first-time sending of messages. No link will be made busy by nodes trying to overcome neighbour-to-neighbor communications problems, so it can be guaranteed that there will be no network congestion.

1.7.3. AbNET Physical Layer: A Hybrid System

In the discussion above, it was assumed that the communications system was exclusively fiber optics. While this is technically feasible, there are considerations that make other communications media worth considering.

First, in a distribution system that is widespread, such as a rural network, the economics do not favor the use of a fiber optics medium. As we showed earlier (Kirkham et al., 1989) UHF radio can be cost competitive at node spacings above about 500 m. Radio transceivers operating at about 950 MHz are costly items, but if their cost is less than the cost of the fiber cable required to cover the distance, their use may be justified. The greatly decreased speed (bandwidth) of the radio system may be a problem, but if the application can be adequately supported even with the reduced performance, radio may be the appropriate choice.

There is nothing in the protocols that would make AbNET unsuitable for a radio-based system. In essence, the fiber optics network considered earlier is a broadcast system, with all nodes receiving all information. A radio system is different in that a receiver could conceivably receive a signal both direct from the base station and repeated by another transceiver. As far as the AbNET protocols are concerned, this would be no different than a node receiving a message over two different fibers. The second message would be ignored.

The second aspect of the physical layer that makes non-fiber systems worth consideration is access to customers' premises. The use of a fiber optics connection for load management, for example, would mean access to the home, as well as modification of existing designs of load management equipment. A hybrid system, (and by this we mean a true hybrid system, that requires no systems engineering to be done by the utility--a marked contrast with today's combined systems) could overcome this disadvantage.

A hybrid that used a high speed network on the distribution feeders (required by feeder automation functions) and low-speed communications on the secondary is attractive. An example of this kind of system (actually UHF radio and DLC) is used in the NetComm project in California (Holte, 1989). We propose to specify AbNET as a hybrid system consisting of fiber optics (or possibly UHF radio) on the distribution feeder and DLC on the secondary.

The use of DLC on the secondary has several advantages. This approach enables the continued use of existing load-control and meter-reading hardware, which in turn implies minimum need to enter the customer's premises for installation. Further, communication into the home can be compatible with other communication applications of the home wiring, for example the PLBus of the Electronic Industries Association (EIA) Home Automation Standard. An EIA PLBus interface at each distribution transformer would allow the utilities to accomplish far more control than simple management of one load. Not only could there be the capability of emergency load-shedding, but a solution to cold load pickup problems would be available.

In cost terms, a hybrid of this kind is certain to be competitive if more than one function is to be performed by the distribution automation system. Feeder automation functions can be implemented competitively (see above) and demand side management can be added at extremely low cost. The small size of the network to be covered by the DLC or home wiring system (typically 3 or 4 houses) greatly eases its communications requirements. Very low power, inexpensive modules operating at about 120 kHz are widely available, and signalling speeds up to 1200 baud are possible. The incremental cost of adding this capability and other functions at the customer level would be minimized.

Complete physical layer specifications will be addressed in our future work.

1.7.4. AbNET Communications Management: The Higher Layers

In computer communications, it is generally assumed that each location is as likely as any other to originate a message. This assumption leads to the need to solve the contention problem for access to the medium. A similar assumption has generally been made in the past in designing communications systems for distribution automation. Because of the generally low data rate of available communications channels, it has been necessary to limit communications traffic by having the RTUs make some intelligent decisions on their own. RTUs are usually designed to operate with software that permits them to originate a transmission only if they detect some drastic change in the data they monitor.

Since the cause of such a change cannot be determined in advance, and might affect several RTUs simultaneously, some kind of medium access protocol must still be used. Collision detection has been used in some demonstrations, and is an automatic feature of the busy signal of telephone-based communications. The problem with this solution is that the response to a collision is always a delay. Whenever a large or widespread change occurs in the distribution system, the communications system becomes overloaded, and information transfer is subject to unpredictable delays.

This is not the case with AbNET. A centralized polling strategy will be used, so that a node (RTU) can transmit only if so directed by the substation unit. In essence this is a token-based system, but only the substation unit can originate a token, and that is addressed to a particular RTU. Before the reader objects that this is slow and inefficient, since it requires that all nodes be scanned before a problem can be discovered, let us point out that we expect a scan of all nodes to occupy less than 20 ms. If the message from one RTU is 1000 bits long, and there are 100 RTUs, a complete scan comprises 10^5 bits. At a data rate of 10^7 b/s, the minimum scan time is 10^{-2} s, or 10 ms. We have doubled this estimate to allow for error correction code overhead, and software execution time.

An advantage of the polling approach is that it allows the substation unit to allocate communications time dynamically. If a problem is suspected in some part of the system, the central unit can concentrate on that area, and relegate the remainder to a lower priority. Whether this would ever be necessary will presumably depend on the application.

It is also apparent that in a polling system, the central unit can simply assign message numbers to its transactions with the RTUs. The first byte of the data could be used as the message identifier, for example. A responding RTU could simply increment the message number. There would then be no need for the RTUs to keep track of the time, or the originator of a message they were handling.

The widespread use of optical bypasses in the network means that the substation unit (the central unit for any given feeder) must be capable of updating its view of the system from time to time. Because of the interconnected nature of the communications network, failure of a single node is unlikely ever to prevent communications to any other location. Such a failure may even be relatively unimportant to the control or monitoring functions for a long time. Nevertheless, the operation of its bypass should not go unnoticed or unrecorded. Because the communications system management software in the central unit can address all nodes, it can be made capable of developing its own system map when it is first turned on.

Suppose each remote unit is equipped with a ROM describing its location, and a few pertinent details of its function. At switch-on, and periodically thereafter, the central unit can poll all the remote units to determine their status and their connectedness. This might be done as frequently as every second. Any changes (in the communications system, not the power system) could be logged automatically. Software could be written so that the operator display could be automatically updated to reflect problems such as the inability to communicate with a particular location.

Even failure of a substation unit can be handled without overloading the communications channel. In the system diagram (Figure 1-1) some nodes were designated "gateways." A gateway is a node that can pass signals from the area of one substation to another if required, so that communications can keep pace with the changing power system configuration that might follow loss of power into a distribution substation.

In short, a communications system can be implemented so as to take advantage of the speed of optical fibers, and incorporate features that will increase its power considerably. These aspects will be explored further in our future work.

1.8. Voice Channel

Finally, following a suggestion made by George Allen of American Electric Power, we propose the addition of at least one voice channel to the fiber optics portion of the communication system. A voice channel may seem at first to be rather wasteful. After all, it will rarely be used. However, it would be very convenient to have a voice channel during initial commissioning and subsequent maintenance of both the distribution automation system and the communications system. Such a channel would allow line-crew personnel installing or calibrating the equipment to talk directly with anyone else on the line, or with colleagues at the distribution substation.

TDM hardware is available that can multiplex voice and data onto the same channel. However, even with the most efficient encoding method, the bit rate required for the voice channel will exceed that of the data. Further, transmission delays must be all but eliminated for a voice channel. It is thus likely that the requirements of the voice channel would dominate the system design in our application. Even so, unless it renders normal system operation impossible, a voice channel should not be ruled out.

It is not clear, however, that the voice channel *should* be added by multiplexing it onto the fiber used for data acquisition and control. A separate fiber, or possibly two fibers, dedicated to voice communications, might make economical sense. The addition of a third and fourth fiber might only add $0.30 to the cost of a meter of cable, or $300 per kilometer, perhaps 10% of its cost. If the cost of adding multiplexing equipment to the remote units exceeded this figure, the separate fiber is justified.

While a careful trade-off study has not yet been performed, it seems likely that additional fibers will be added. In each remote unit the voice fibers can either be terminated at an optical bypass as part of a special connector, or passed through low cost repeater equipment that could be tapped into as needed. The fibers need only be accessed at a few locations at any time, so they do not have to be managed as part of the general data acquisition and control system. A low-cost party-line system using pulse-code modulation (PCM) would be perfectly adequate.

1.9. AbNET Summary

AbNET is a dual-hybrid network, comprised of a fiber-optics-and-power-line-carrier digital data system in which every node (or RTU) is also a repeater, and a passive fiber optics PCM voice channel sharing the fiber cable. The nodes are interconnected by fiber optics links routed along the distribution system. The fixed installation of the voice channel is intended to be passive, the opto-electronics and the electronics being carried by the user and inserted into the system as needed. The digital data system, on the other hand, can be put in the framework of the ISO OSI model, as shown in Figure 1-6.

APPLICATION	DISTRIBUTION AUTOMATION	CENTRALIZED, LOCATED AT DISTRIBUTION SUBSTATION
PRESENTATION	ADDRESS TRANSLATION	
SESSION	POLLING	
TRANSPORT	ERROR CORRECTION CODES	
NETWORK	FLOODING ANTIBODY MESSAGE REMOVAL	DECENTRALIZED, DISTRIBUTED AT ALL RTU'S
LINK	SIMILAR TO IEEE 802.5 BUT NO FLOW CONTROL	
PHYSICAL	FIBER OPTICS AND POWER LINE CARRIER	

Figure 1-6. Main features of AbNET in OSI framework

Distribution automation is shown in Figure 1-6 as the applications layer. Descriptions of the various functions that comprise distribution automation need not be repeated here. Suffice it to say that AbNET was developed explicitly to satisfy the communications requirements of data acquisition and control for distribution automation. The brief descriptions that follow show how AbNET meets these requirements, and supports distribution automation.

The principal function of the presentation layer is to perform the conversion of data descriptions (in application programs) into addresses that the lower layers can recognize.

The session layer establishes transport connections between the central unit and the remote nodes on a polling basis. Centralized information about network performance can only be made available to an applications program through the operation of this layer. The session layer therefore retains the ability to establish (and test) connections between adjacent nodes.

The transport layer is centralized at the distribution substation. It normally operates through the local network in support of the data acquisition and control functions. Under unusual circumstances, it can choose to use the local network for access to the area served by another substation. Message integrity is improved by error detection and correction codes, but the choice of Hamming, Reed-Solomon or BCH has not yet been made.

The network layer is a decentralized message-based, store-and-forward system. Congestion in the network is controlled by an antibody-like algorithm, that allows a node to repeat a message once and only once. Routing is accomplished by network flooding.

The data link layer incorporates the field descriptions and framing definitions of IEEE 802.5 as far as possible. There is no handshaking for flow control, and no request for retransmission in the event of an error.

The physical layer is a fiber optics based hybrid. Exact details are not fixed at the time this is written, but it is very likely that multimode fiber will be used, transmitting optics will be based on LEDs operating in the near infrared, the bit rate will be between 10 and 15 MHz, and differential Manchester coding will be used. Access to locations on the power line secondary will be by means of low-power high-frequency power line carrier communications.

1.10. Concluding Remarks

In the discussion above, mention was made of the anticipated data rate of the fiber optics communications system. Of course, such a scheme could be operated at a rate barely great enough to meet the estimated requirements, 1 kbit/s. However, the performance of even a minimum fiber channel will provide excess capacity of three orders of magnitude over this data rate, at no additional cost. To appreciate what that really implies, consider that a jet airplane is only two orders of magnitude faster than walking!

The combination of features in AbNET has several advantages.

> The flooding approach provides maximum communications reliability, even in the event of partial system failure. It also guarantees that messages are transmitted over the shortest possible route, minimizing transmission delay.
>
> Polling by the substation unit solves contention issues, and allows the distribution substation to allocate priorities dynamically if necessary.
>
> The antibody algorithm and the use of error-correction coding obviate the need for node-node handshaking. This means that node processor time and communications link time are not affected by the details of communications problems.
>
> Adoption of the definitions and formats of IEEE 802.5 mean that a good deal of the detailed design of the system is already standardized.
>
> The fiber optic cable used is all-dielectric, sometimes a useful feature in the neighborhood of power lines. Further, the communications channel is unaffected by electromagnetic interference from the power line or other source.
>
> The bit rate of the fiber optics system is high enough that a considerable spare communications capacity exists, allowing for the development of new functions in the future. It is not so high that costs are increased by the need for single-mode or high-performance components.

Centralizing the higher layers of the network and decentralizing the lower layers allows for the efficient development of applications, and should stimulate competition in both hardware and software.

The last two items listed above are two aspects of the AbNET approach that may well change considerably the way in which system designers view the function of distribution automation. Fiber optics will remove the bandwidth constraint that has limited planners and designers up to the present time. In addition to accomplishing the original goals of making distribution automation practical, there may well be new applications for the communications system that will take directions far from the original motivation for distribution automation. Imagination will be the only limitation. For example, although the utility is statutorily obliged only to supply power within certain voltage and frequency limits, so many users have come to rely on the power company for timekeeping that separate time signals may some day be sent over the communications link to reset consumers' solid-state clocks after a power failure. The history of advances in other technologies strongly suggests that other useful and unanticipated applications will in fact arise.

Further, AbNET marks the end of an era dominated by vertical integration. The utility wishing to implement distribution automation will no longer be dependent on a single supplier for his hardware, software and communications system, and for after-sales support. The AbNET approach isolates the distribution automation functions from the communications system to the maximum extent. We envision the AbNET system playing the role in distribution automation that the IBM-PC and its DOS played in the world of small computers. Since the control and monitoring functions are centralized at the distribution substation, it will be possible for competition to develop in the market for distribution automation software to be used there. In this, there is a resemblance to the proliferation of, say, word processing software in the PC world.

Extending the parallel with PCs, it is anticipated that the AbNET hardware and software, built to a published performance standard, could ultimately be available from a number of sources.

Before any of this can happen, of course, it is necessary for the system to become established as a *de facto* standard in the industry. We are presently seeking ways to promote these economic goals. If we are successful, the demonstration of a fiber-based distribution automation system may soon herald the beginning of a new era in utility communications and control.

2. AN IMPROVED FM LINK

2.1. Introduction

The use of optically-supplied energy as the sole source of power is well known in the communication satellite industry. At distances from the sun up to about the orbit of the earth, the intensity of solar radiation in space is such that a reasonably-sized photovoltaic array can power a significant electrical load. Our application is more down-to-earth, but has the same general principle: optical power can be used to furnish energy to a location that is otherwise rather inaccessible.

In a power system application, optical powering allows the instrumentation designer to create a sensor that has the advantages of conventional electronics, while having the external attributes of being optical. The only connections to the device are optical fibers, that can easily be contained inside a composite insulator, so that the sensor could be operated at line potential. There is no need for an auxiliary power supply, or the complex temperature compensation sometimes needed with optical measurements.

In an earlier report, we described a practical link that transmitted optical power to a remote transducer, and data back. Sufficient spare power was available to energize an electronic measurement system.

The prototype link had moderate bandwidth (1 kHz), accuracy (1%), and dynamic range (>60 dB). While these are analog specifications, the link used a frequency modulated optical pulse train to retain noise immunity and insensitivity to changes in the fiber loss characteristics. The approach could be used with any of the usual electrical or electronic measurements: current transformers, strain gauges, thermocouples, etc. As an example, we presented results showing current measurement by means of a linear coupler.

However, in terms of bandwidth and dynamic range, this performance was achieved through the use of a post-detector filter. In our prototype, a six-pole Bessel filter was used, to provide a reasonably flat frequency response over the passband of interest. This kind of filtering requires careful choosing of component values, and necessarily incurs a penalty in the phase response of the system.

In the present report, we show how these difficulties can be avoided by means of a different approach to the process of demodulation or modulation of the signal in the optical link. The earlier advantages of simplicity in the probe, and a sufficiently low power consumption to allow optical powering are retained.

2.2. Optically Powered Links

There has been only a small number of fiber (as opposed to solar) powered optical links reported in the literature. The earliest seems to have been for telephones: a system was described by DeLoach, Miller and Kaufman at Bell Labs. in 1978. In 1981, an approach to optically powering an implant in a human body was described by DeLoach and Gordon. In 1984, a temperature measurement system was described by Ohte, Akiyama and Ohno; and McGlade and Jones described an optically-powered force sensor. Perhaps because the amount of electrical energy available at the load was small, this method of energizing electronics did not generate widespread commercial interest. Nevertheless, it would seem perfectly suited to power system applications where the advantages of electrical isolation and immunity to noise are important.

There have been other, more recent, examples of optical powering of electronic systems. A computer-controlled sensor network was described by Hall in 1986. In 1987, a pressure transducer was described by Schweizer, Neveux and Ostrowsky. However, there seem to have been no power-system applications until a 1989 paper by Adolfsson, Einvall, Lindberg, Samuelsson, Ahlgren and Edlund that described an optical CT using this principle.

The principle of using optical power for sensing is very straightforward. A representative example is shown in Figure 2-1. Light energy, typically from a solid-state laser, is coupled into a fiber at some convenient place, ie, where power is available. The other end of the fiber is brought to where the power is needed, and the light energy emerging from the fiber is coupled into some optical-to-electrical convertor, typically a photovoltaic device. The output voltage is a function of the configuration of the receiving convertor (are several cells in series?) and the temperature (photovoltaic convertors generate less voltage at higher temperatures). The current available is determined by the amount of light, the efficiency with which this light can be coupled to the convertor (is the illumination uniform, is there light missing the device, are there surface reflections?) and the conversion efficiency of the cell (does the material bandgap match the incident wavelength, are the device doping and structure appropriate?). Some of these issues have been addressed in an earlier report (Kirkham, Johnston, Lutes, Daud and Hyland, 1984).

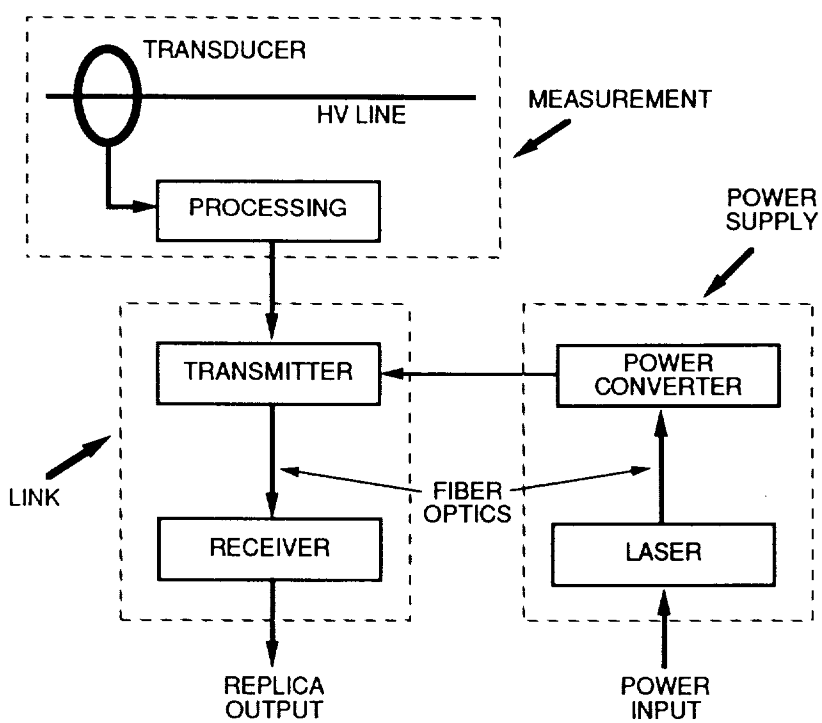


Figure 2-1. Representative optically powered sensor

Optical coupling efficiencies depend on the characteristics of the laser, the fiber and the convertor. The physics of the situation is such that there are inevitably losses; it would be remarkable if half the light output of the laser were available at the photovoltaic device. It would be remarkable, too, if half this energy could be converted into electricity. In our prototype diode arrays, overall efficiency (electrical-to-electrical) was in the order of 5%. While the system of Adolfsson *et al.* is claimed to have rather higher efficiency, their device is nevertheless operated at an extremely low power level, through the use of custom integrated circuits.

The small amount of available electrical power is a severe constraint. Even though low-power integrated circuits are available, sensor power must be carefully budgeted. And however frugal the system designer can be with signal processing power, there is another constraint in terms of the optical power budget of the link: it takes a certain amount of optical power to cause the receiver photodetector to work.

An optical link designed to have minimum (average) power consumption will use very short, intense optical pulses. The optical power level is set by the receiver sensitivity, and the need for a certain extra power (the power margin) to accomodate temperature effects, connectors, and additional unpredicted losses in the link. The pulses cannot be made arbitrarily short, either. Apart from the difficulty of producing extremely short electrical and optical pulses, as pulse width is decreased the required power level at the receiver is increased. These limitations mean that, once a pulse generating system and an optical receiver system are chosen, the only parameter that the system designer can use to minimize power consumption is the link frequency.

It is because of these considerations that the optical links used in our recent field meter work (for example, Kirkham and Johnston, 1988) operate with carrier frequencies in the order of 5-10 kHz, rather than the ~100 kHz of our original (battery powered) unit.

However, the use of a low carrier frequency impacts the bandwidth of the system. Nominally, the process of demodulation is the extraction of the modulation information from the modulated carrier. Whatever demodulation process is used, a certain amount of the carrier is bound to remain. When the carrier frequency is decreased, unless the bandwidth is also decreased, the problem of separating this residual carrier from the desired output becomes more difficult.[1]

Normally, the separation process is performed by frequency-selective filters in the analog domain. While this approach is practicable, the drawbacks of complex

[1] An analogous situation arises in reconstructing analog waveforms from the information on a digital Compact Disc. An elegant solution in many modern CD players is to use the technique of oversampling. This can be explained as follows. Imagine the output waveform to be reconstructed by means of a sample-hold system, or staircase generator. Since the sample rate on CDs is about 44 kHz, the problem is to remove the remnant 44-kHz component without affecting the desired audio at, say, 20 kHz. By interpolating between adjacent values of the signal, additional steps can be created in the reconstructed output. The apparent sampling rate can be increased by a factor of 2, 4 or even 8 times. This means that the undesired residual can be placed as much as 16 times higher in frequency than the desired output. A very simple filter can accomplish the task of separating such signals.

This is not an option open to us, since we are dealing with real-time signals, not stored data. The delay involved in waiting for samples for interpolation is unacceptable.

filtering and modified phase response make it worthwhile to investigate other solutions.

2.3. Phase-locked Staircase Generator

At most "ordinary" frequencies, a phase-locked loop can be used as a demodulator of FM signals. We use this as a starting point.

A phase-locked loop consists of 3 essential elements[2]: a phase detector or comparator, which compares the phase of an incoming signal with that of a locally generated signal; a loop filter, which smooths the phase comparator output to produce an error voltage; and a voltage controlled oscillator, whose frequency is a function of the applied error voltage. They are typically arranged as in Figure 2-2.

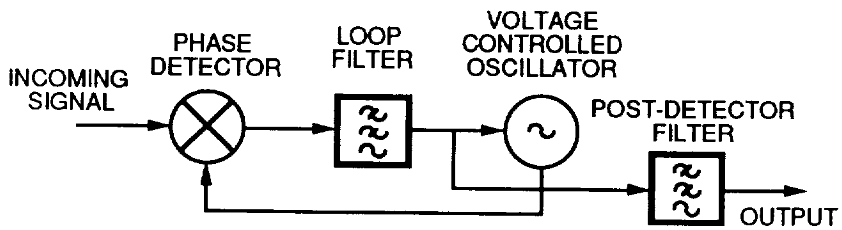


Figure 2-2. Typical phase-locked loop

The phase comparator and loop filter determine the lock range, the capture range and the noise performance of the loop. If the incoming signal has a large signal-to-noise ratio, one is less concerned about noise performance. When used as an FM demodulator, the loop filter has to be a compromise between flat frequency response and rejection of the unwanted phase comparator products. As the spectral frequency of the incoming modulation approaches the carrier frequency, this compromise becomes difficult to resolve.

As an illustration of the nature of the problem, note that with the typical multiplying-type phase detector, the lowest frequency component of the phase detector output is at twice the carrier frequency, f_o. However, the amplitude of this component is large, regardless of the amplitude of the modulating signal, because this kind of phase detector switches between two extreme states. Typically, the signal applied to the low-pass loop filter is a square wave with peak amplitude equal to the local power supply voltage.

Stability considerations, and simplicity of calculation, dictate the use of simple low-pass or lag-lead loop filters. The filter of a wideband loop removes

[2]There is also the possibility of nonessential elements, such as frequency dividers (which can be used in the forward or the feedback parts of the loop), and presettable counters, which can be similarly applied. These components do not affect the present discussion.

little of the energy of the signal at $2f_0$. For an undistorted output, a post-detector filter of considerable quality is required. Such a filter is shown in Figure 2-2.

The first of our two revisions to the optical link greatly reduces the filter requirements. By replacing the phase detector and loop filter with a sample-hold system, or staircase generator, the ripple component of the phase detector output is greatly reduced. With this modified loop, shown in Figure 2-3, the total ripple energy is dependent on the signal amplitude, and is largely in the high harmonics.

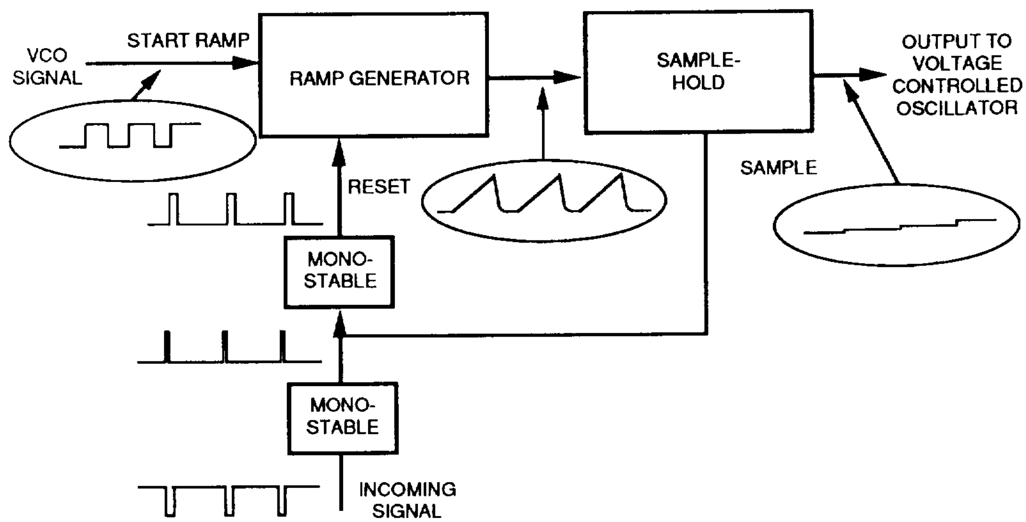


Figure 2-3. Block diagram of sample-hold system

Operation is as follows. Suppose that a reset pulse has just set the ramp generator to zero. The system is static until a signal comes from the VCO. This starts the ramp generator, which can be a simple op-amp integrator. A short time later, a pulse comes from the incoming signal line. This triggers a monostable oscillator, which produces a short pulse that tells the sample-hold circuit to hold. The voltage that appears on the output of the sample-hold is thus a measure of the time difference between the VCO and the incoming signal. The trailing edge of the monostable triggers a second monostable, arranged to reset the ramp generator. The second monostable serves to delay the reset until the sample-hold has acquired a new value. The sequence repeats.

The timing diagram (Figure 2-4) shows the relationship between the various signals. Starting at the bottom of the diagram, the incoming signal causes the first monostable to generate the SAMPLE pulse. This puts the current value of the ramp generator on the output of the sample-hold circuit. The trailing edge of the SAMPLE pulse starts the RESET pulse, which causes the ramp generator to discharge. The ramp circuit starts to charge in response to the VCO signal. For a decreasing frequency of the incoming signal, the output is the familiar staircase waveform.

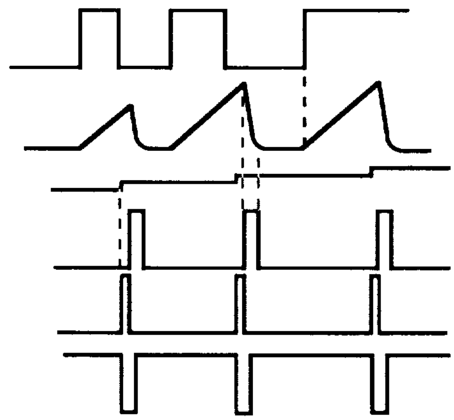


Figure 2-4. Timing diagram

Because the circuit is inside a phase-locked loop, the time difference between the incoming signal and the VCO is a measure of the phase difference. The ramp and sample-hold circuits together perform the function of phase detector.

2.3.1. Higher order loop

It is possible to look at this another way. The interval between the pulses of incoming data in a frequency modulated data link contains information about what is going on at the transmitter. The interval between any two adjacent pulses can therefore, in principle, be used to *estimate* the state of the transmitter. Some of the techniques we have used to implement the new demodulator system are borrowed from the world of state estimation.

In essence, the device attempts to duplicate in the demodulator the events going on in the modulator. The interval between pulses in the telemetry link is used as an estimate of the voltage on the control terminal of the VCO. A voltage derived from this estimate is applied to the demodulator VCO, which therefore mimics the transmitter VCO. The accuracy of reproduction is controlled by the fact that the loop is closed: any deviation represents an error that will generate a correcting signal in the phase-locked loop.

The simplest way of estimating what is going on in the transmitter is to assume that the control voltage applied to the VCO did not change during the period between output pulses. This is what was described above; it results in a loop that approximates the performance of a conventional first-order loop, although the estimate of the modulating signal is known as a zero-order estimate. (From now on, we must distinguish between the order of the estimator and the order of the loop.) If the modulating signal is constant there is no error. If the signal is changing at a constant rate, there is an error due to the steplike nature of the approximation. This can be seen in Figure 2-4.

If information about the way the period changes from interval to interval is incorporated, a first-order estimate results. In a first-order estimator, the assumption is made that the *change* during the period is constant. The error is consists of small ramp segments, rather than small constant segments. The design of a loop using such an estimator is discussed next.

2.3.2. First-order estimator: loop implementation

In the general state estimation scheme, the first-order estimate can be obtained by incorporating information about the *difference* between adjacent samples of the data. For example, suppose that between one sample and the next, the value is found to have increased by 10%. The zero-order system would assume that the last measured sample is the best estimate of the next. The output of the system is constant until the next sample arrives. The first-order estimator would assume instead that there will be a further 10% increase in value during the next interval. The output is a ramp, starting at a value that represents the last measurement, and calculated to increase the output by 10% during the next period. The hardware for this is clearly going to be a little more complex than for the zero-order system.

Our first-order estimator functions as follows. Control signals (short pulses) are derived, as before, from the incoming signal by means of simple delay circuits. The details of this are not shown in the block diagram, Figure 2-5. A zero-order system is implemented, as before, by means of a ramp generator and a sample-hold circuit. The ramp generator of this part of the system now contains circuitry to enable it to hold the output constant for a brief interval before it is reset.

To make the estimator first-order, additional hardware computes the difference between adjacent intervals. An analog difference circuit continuously monitors the difference between the output of the zero-order hardware just described and the input to its sample-hold circuit, see Figure 2-5. At the instant that the zero-order sample-hold circuit is instructed to sample, this difference represents the change in the signal between two adjacent intervals. This value is therefore sampled just before the sample in the zero-order part of the system.

A second ramp circuit is used to generate a ramp whose slope depends on the output of the difference sample-hold circuit. The direction of the ramp depends on the sign of the difference, and the steepness of the ramp is controlled by the magnitude of the difference signal. These waveforms are shown in Figure 2-6 for a high modulating frequency (few samples per cycle).

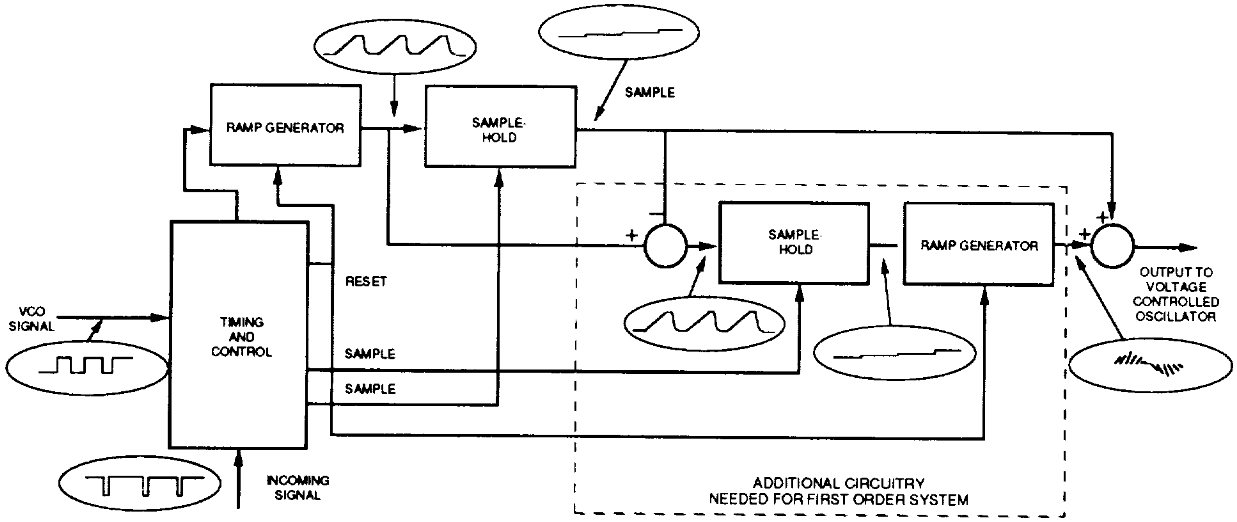


Figure 2-5. First-order estimator loop

The first-order estimate is obtained by adding the zero-order estimate and the difference-generated ramp, ie, by adding the two signals shown in Figure 2-6.

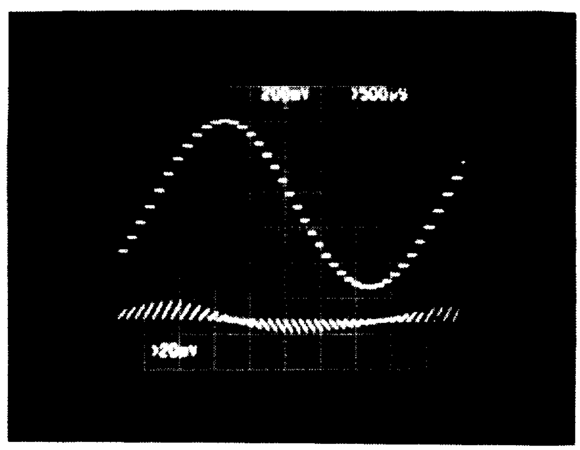


Figure 2-6. First-order estimator waveforms

2.4. Zero-order estimator loop - measured performance

The simplest implementation of our loop uses the zero-order estimator, ie, one that has no derivative compensation. This version has, nevertheless, significant performance advantages over conventional loops. To demonstrate this, we arranged loops of various kinds, conventional as well as sample-hold, as demodulators of a low-frequency FM carrier. All loops were fed the same signal. The results stress the low noise content of the demodulated signal from our loop.

There are normally considered to be 2 kinds of phase-locked loop, distinguished by their phase detector. A loop using the original multiplying-type phase detector is known as a Type I loop. This phase detector has the characteristic that, with lock at the VCO free-running frequency, the output waveform has unity mark:space ratio. After filtering, this signal averages to zero, satisfying the condition for lock at the free-running frequency. There is a 90° phase angle between the input signal and the VCO. At other frequencies in the lock range, there is a different phase shift between the input signal and the VCO signal. A non-zero error voltage is generated because the mark:space ratio changes.

In most implementations, the VCO waveform is a square wave. The phase detector output waveform depends on the signal waveform. If the signal is sinusoidal, the filtered phase detector output is a sinusoidal function of the phase angle. If the input signal is a square wave, the filtered phase detector output is a linear function of the phase displacement. These characteristics are well known. (See, for example Gardner, 1966.)

What is sometimes overlooked is that the *unfiltered* phase detector output is a large amplitude signal. In applications where the input signal is amplitude limited before the phase detector, the output waveform is a square wave whose instantaneous amplitude is, for all practical purposes, equal to the system power supply voltage.

The other conventional phase detector (Type II) is based on the detection of zero crossings in the input and VCO signals. When locked at the free-running frequency, this kind of loop has 0° phase angle between the signal and the VCO. Their zero crossings are coincident. The phase detector output is zero. If the VCO should drift slightly, the zero crossing of one signal will move ahead of the other. Depending on the relative phase, the phase detector produces an output impulse of positive or negative polarity. The filtered pulse acts so as to correct the VCO frequency. If the drift (or frequency shift) is slight, these pulses are produced only occasionally. If the loop is operating further from the free-running frequency, these pulses may be produced every cycle.

By extension of this naming convention, we designate our phase-locked loop a Type III loop in the following discussion.

Figures 2-7 through 2-9 compare the output of the three loops for an input modulation of 20 mV. (This is about in the center of the usable range of input signals.) In Figure 2-7(a) the loop error voltage (ie, the output of the loop when used as a demodulator) for a Type I loop is shown. In operation, this signal would be filtered to remove the modulation content, in this case at 80 Hz. Figure 2-7(a) shows that the loop output is predominantly a square wave at the carrier frequency (about 7000 Hz), effectively pulse width modulated at 80 Hz. The signal spectrum is shown in Figure 2-7(b). Here it can be seen that the carrier is over 30 dB higher than the signal that is desired. This means that the filtering that is applied must cope with a "noise" over 30 times higher than the desired "signal".

Figures 2-7(c) and (d) show the performance of the Type II loop. In (c) the loop error voltage is shown. This is clearly an improvement over the Type I, but the improvement is not as great as it seems at first sight. The desired signal is not the large sine wave visible in the figure, it is the much smaller variation seen in the shift of the line near the center of the picture. The more obvious pulses, which appear to have the proper shape, are actually pulses with most of their energy at high frequency. The fundamental of these pulses is visible in the spectrum of Figure 2-7(d). It can be seen that the noise component is still almost equal in magnitude to the signal.

The loop output in our Type III loop is shown in Figure 2-7(e). The output is made up of short segments--but apart from this no noise is visible. A clean sine wave of 20 mV seems to have been reconstructed. It would seem that no further filtering is required. This impression is confirmed by the spectrum of Figure 2-7(f). The carrier component is now a full 40 dB (a factor of 100) below the signal. It is important to understand that this signal has not been filtered at all. The advantage of the Type III loop is that very little filtering is required to recover a very clean signal.

Figure 2-8 shows the performance of the three loops with a larger signal. The signal is 200 mV in this case. In Figure 2-8(a) and (b) the Type I loop is seen. The loop error voltage seems to be about the same as before, a square wave of over 5-V peak amplitude. The spectrum of (b) shows how the loop has changed. The peak amplitude of the carrier is now only about 10 dB (a factor of 3) above the signal, but the total carrier energy is now spread over a wider part of the spectrum. A component at half the carrier frequency has now appeared. This small noise is thought due to the particular way we implemented the loop.

The Type II loop output, in Figure 2-8(c) shows clearly that the large pulses are not the desired output. They have saturated in the Figure. The change in level near the middle of the trace is now more obvious: this is the signal. Figure 2-8(d) shows the spectrum. The signal is now about 10 dB above the peak of the carrier, but the carrier has spread considerably. Its energy content is still large compared to the desired signal.

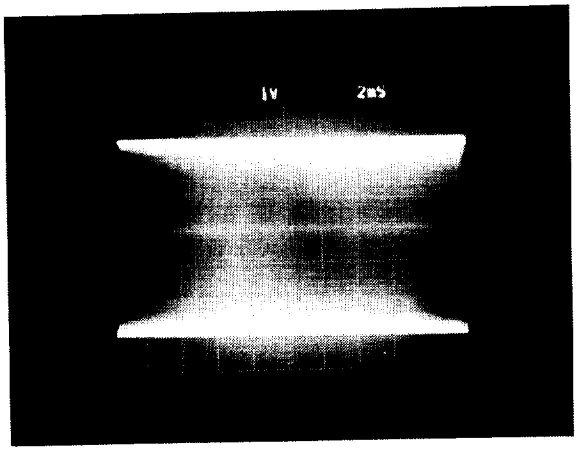
The output of the Type III loop has changed little. It is still comprised of small segments, and the same number of segments seem to be present. The spectrum shows that the signal is now more than 40 dB above the carrier. As in the other loops, the carrier has spread because of the depth of modulation.

Loop performance with a small signal is shown in Figure 2-9, where the modulating signal is 20 mV. As expected, the Type I loop output (Figure 2-9(a)) is a square wave, with amplitude set by the power supply voltage. The spectrum (b) shows that the desired signal is almost 60 dB (a factor of 1000) below the carrier level. This presents a considerable problem to the filter stages that would be required to recover the signal.

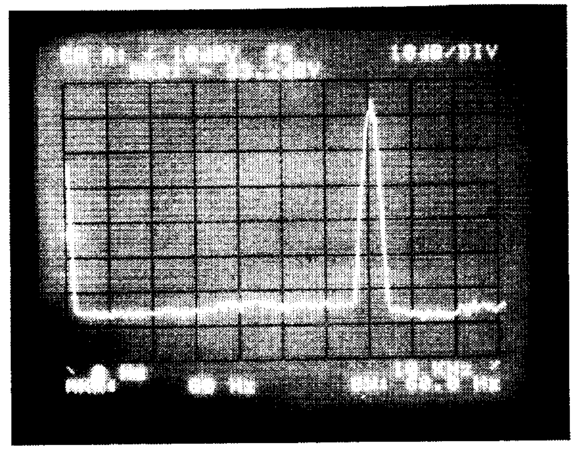
The noise spikes are the dominant feature of the Type II output, shown in Figure 2-9(c). The spectrum (d) shows that they are equal to or larger than the signal. A half-carrier-frequency component is also visible, making the filtering problem worse.

Some noise is visible in the Type III output (e), both in terms of high frequency noise spikes, and random level fluctuations. The spectrum (f) shows that the signal is still almost 40 dB above the carrier, however. Cleaning this output up by filtering it would not provide much of a challenge.

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

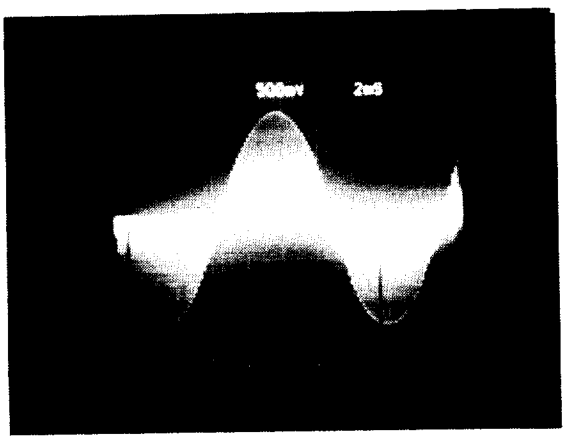


(a) Loop output signal

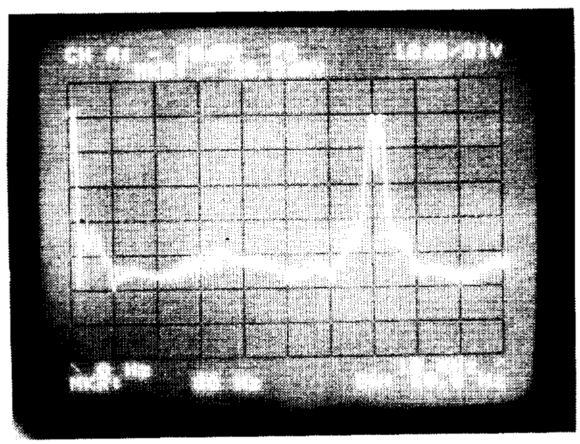


(b) Output signal spectrum

Figure 2-7 (a) and (b). Type I phase detector performance, V_{in} = 20 mV

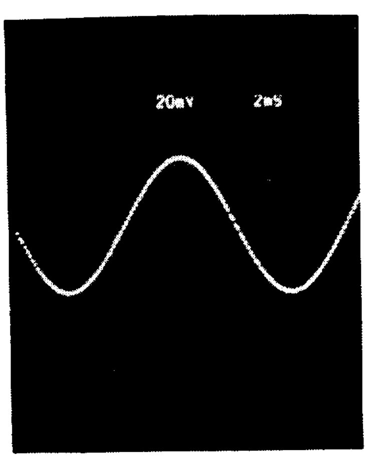


(c) Loop output signal

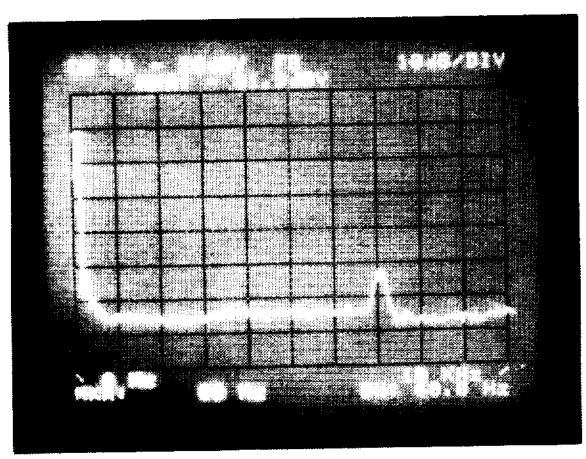


(d) Output signal spectrum

Figure 2-7 (c) and (d). Type II phase detector performance, V_{in} = 20 mV

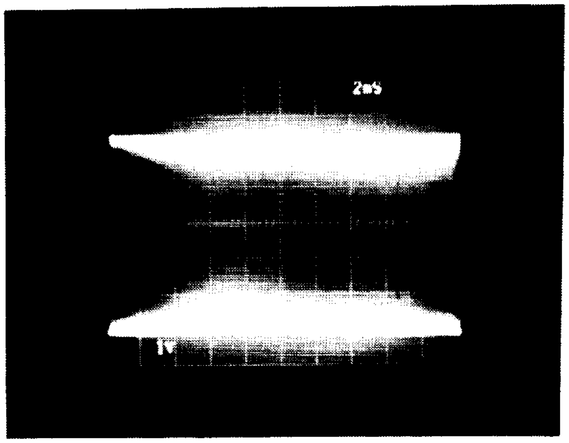


(e) Loop output signal

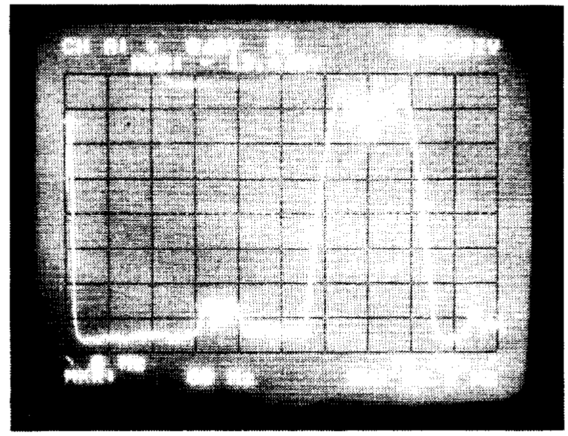


(f) Output signal spectrum

Figure 2-7 (e) and (f). Type III phase detector performance, V_{in} = 20 mV

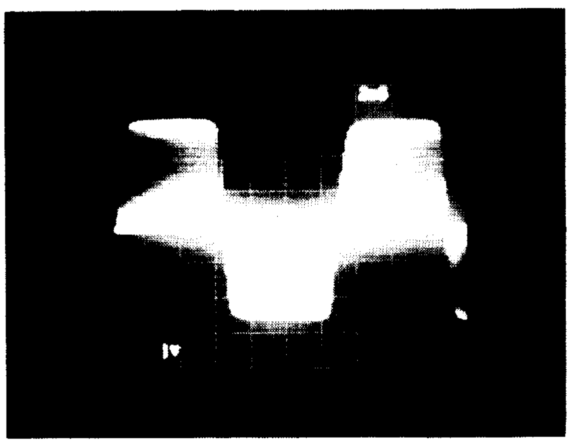


(a) Loop output signal

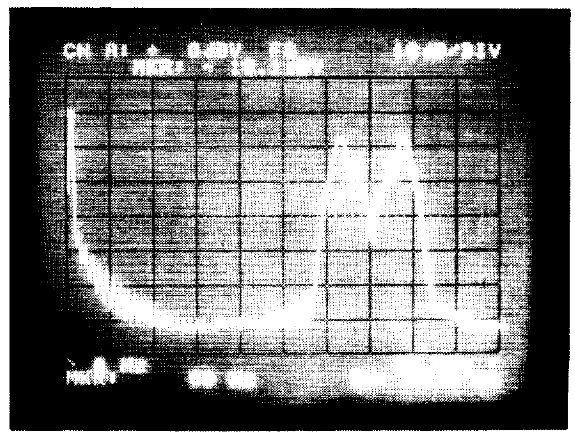


(b) Output signal spectrum

Figure 2-8 (a) and (b). Type I phase detector performance, V_{in} = 200 mV

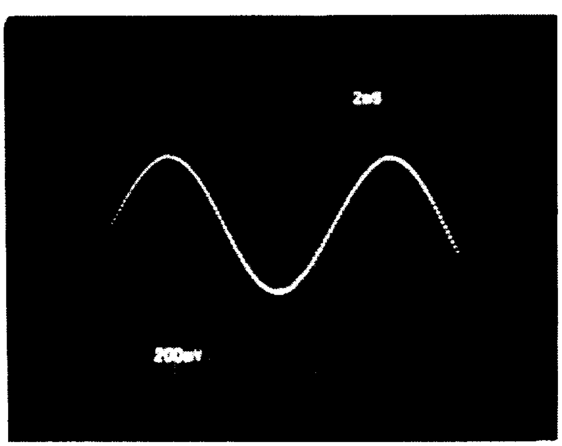


(c) Loop output signal

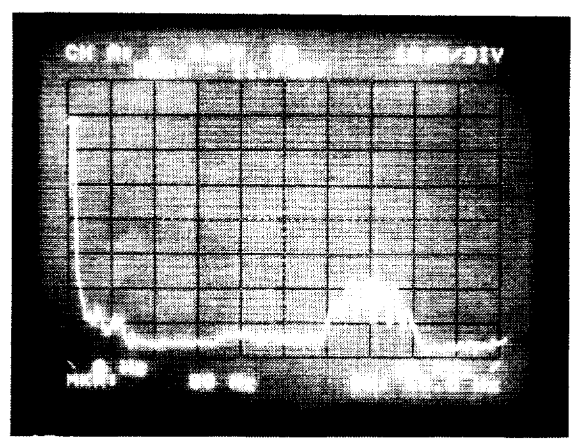


(d) Output signal spectrum

Figure 2-8 (c) and (d). Type II phase detector performance, V_{in} = 200 mV



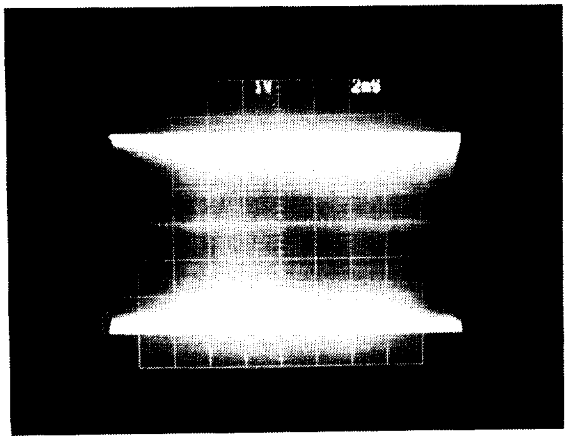
(e) Loop output signal



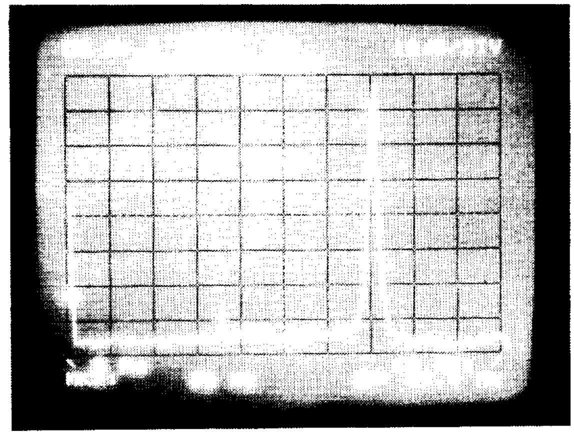
(f) Output signal spectrum

Figure 2-8 (e) and (f). Type III phase detector performance, V_{in} = 200 mV

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

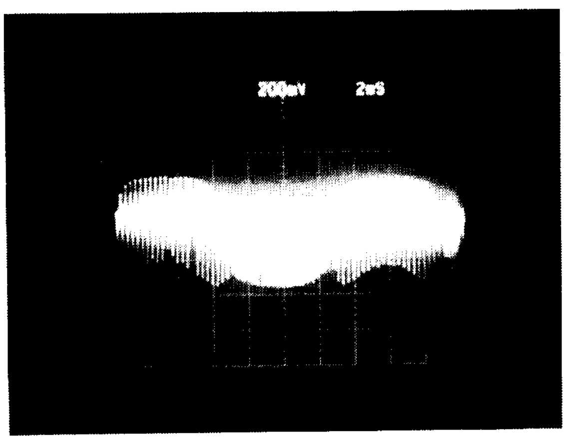


(a) Loop output signal

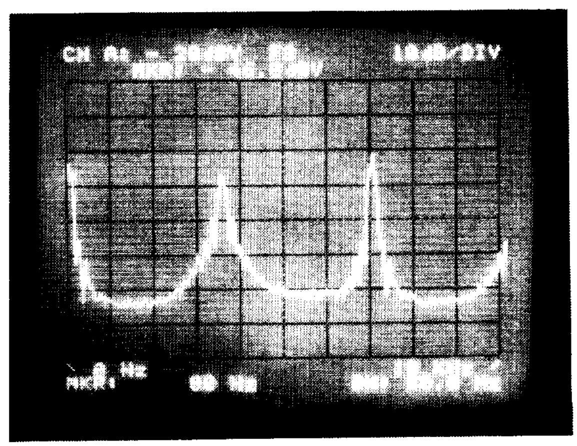


(b) Output signal spectrum

Figure 2-9 (a) and (b). Type I phase detector performance, V_{in} = 2 mV

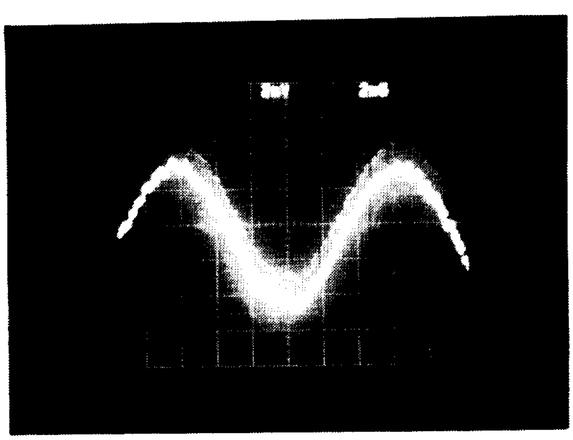


(c) Loop output signal

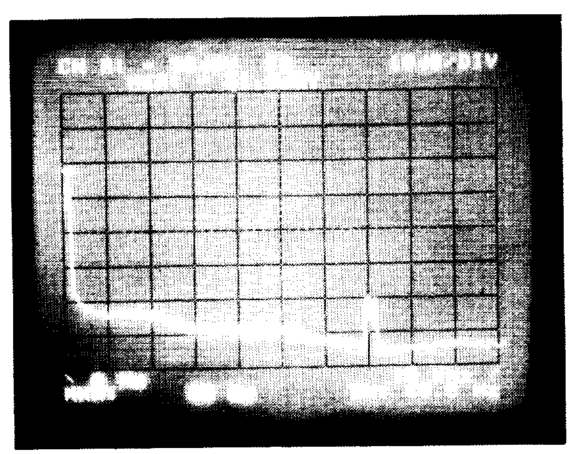


(d) Output signal spectrum

Figure 2-9 (c) and (d). Type II phase detector performance, V_{in} = 2 mV



(e) Loop output signal

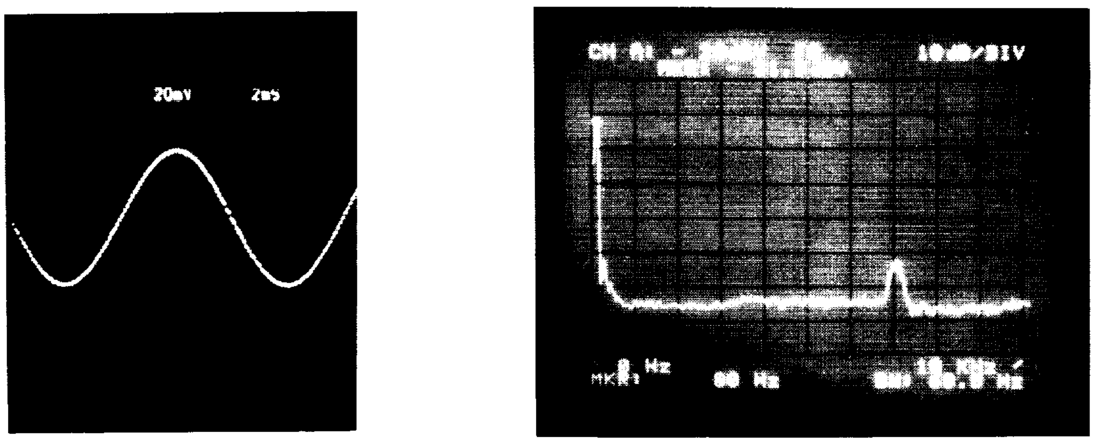


(f) Output signal spectrum

Figure 2-9 (e) and (f). Type III phase detector performance, V_{in} = 2 mV

2.5. First-order estimator - measured loop performance

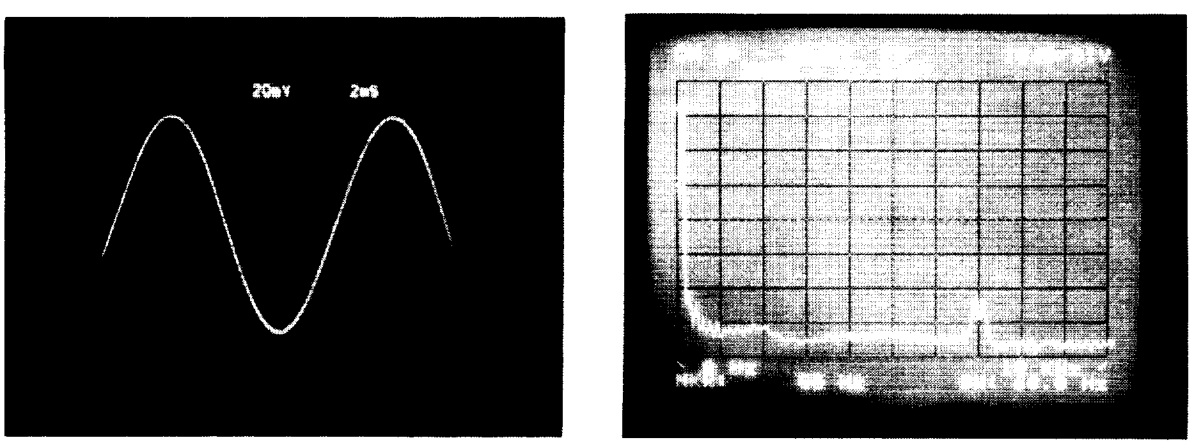
As may be expected, the output of the loop with a first-order estimator "looks" a lot better than the output of the zero-order version. The discontinuities in the waveform are much smaller because of the ramplike nature of the estimate. On the spectrum analyzer, some of the expected improvement disappears. There are two reasons for this. First, the first-order estimator requires more digital switching signals, and these fast, short pulses tend to feed through into some of the analog circuitry. Second, the difference signal is a small value obtained by subtracting one large signal from another. Such a process is inherently noisy. Nevertheless, the carrier rejection properties of the loop are measurably better than those obtained with the zero-order estimator. Figures 2-10 through 2-12 compare loops made with zero-order and first-order estimators.



(a) Loop output signal

(b) Output signal spectrum

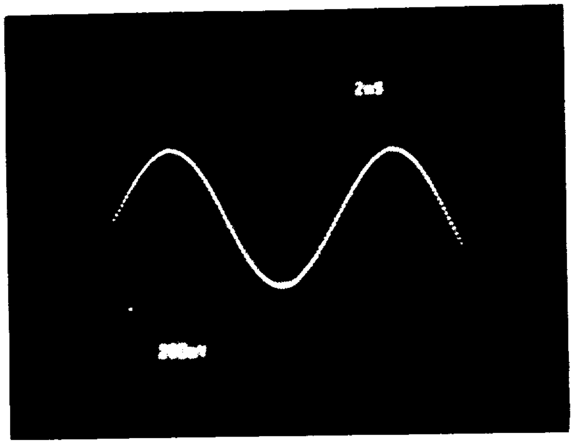
Figure 2-10 (a) and (b). Zero-order estimator, V_{in} = 20 mV



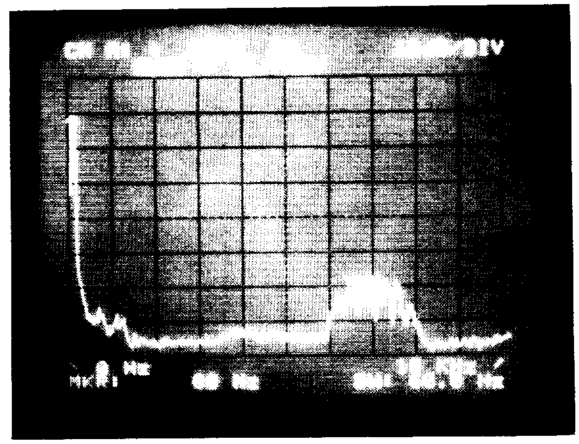
(c) Loop output signal

(d) Output signal spectrum

Figure 2-10 (c) and (d). First-order estimator, V_{in} = 20 mV

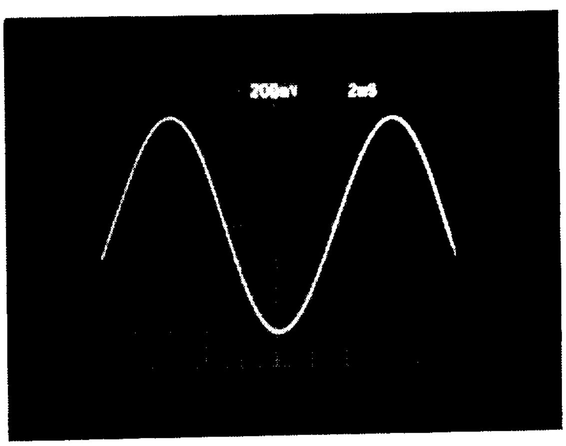


(a) Loop output signal

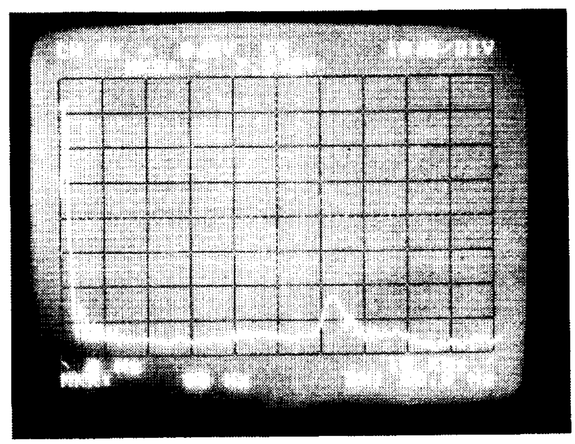


(b) Output signal spectrum

Figure 2-11 (a) and (b). Zero-order estimator, V_{in} = 200 mV

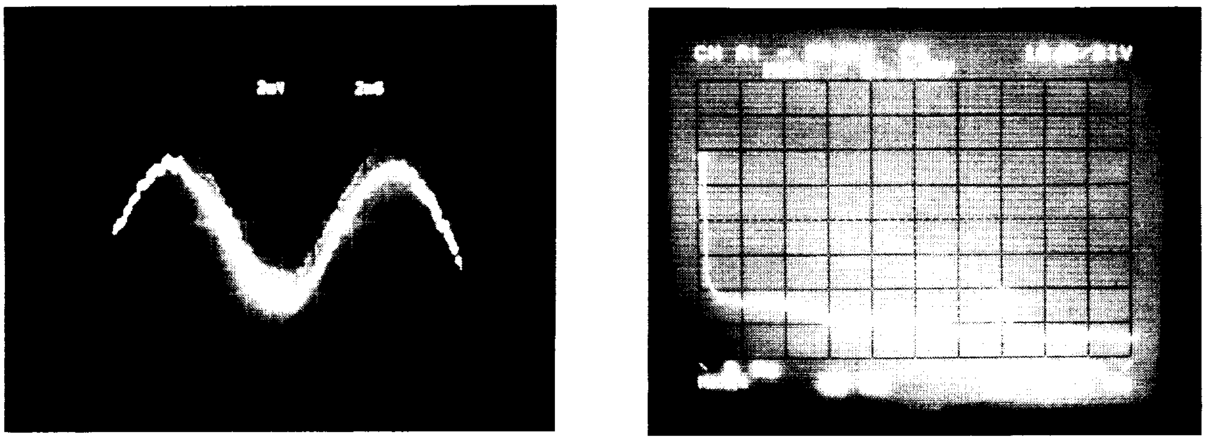


(c) Loop output signal



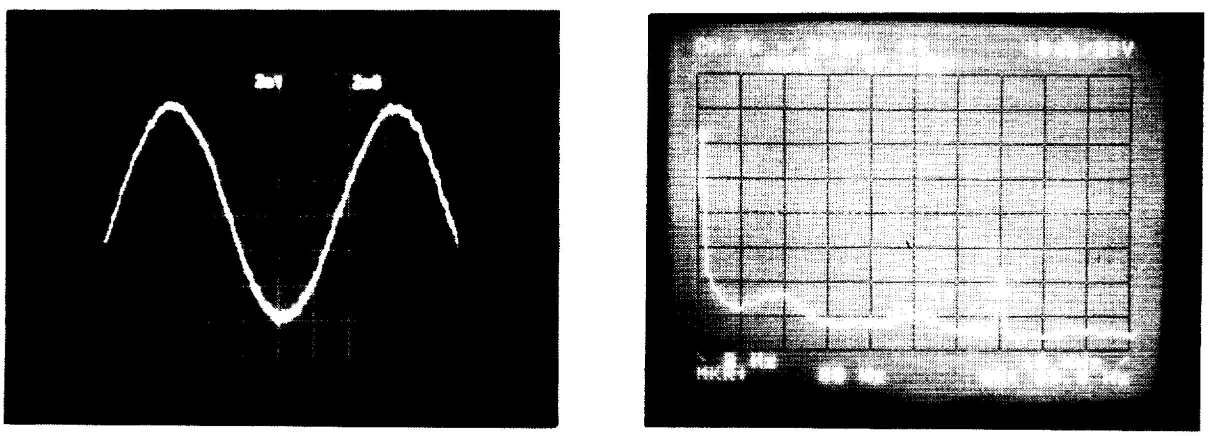
(d) Output signal spectrum

Figure 2-11 (c) and (d). First-order estimator, V_{in} = 200 mV



(a) Loop output signal (b) Output signal spectrum

Figure 2-12 (a) and (b). Zero-order estimator, V_{in} = 2 mV



(c) Loop output signal (d) Output signal spectrum

Figure 2-12 (c) and (d). First-order estimator, V_{in} = 2 mV

There is no doubt that the output of the loop is a remarkable reconstruction of the modulation. It should be remembered that the waveforms shown here are not filtered at all. Not only is the additional circuitry of an analog filter not required, but the unavoidable effects of such a filter on the frequency response and the phase response of the loop are avoided. These topics are discussed further below.

2.6. Additional observations

The phase-locked loop implemented here contains both linear and nonlinear elements. Analysis of the loop performance is therefore somewhat difficult. However, its behavior can be considered by comparing it to a conventional PLL, with some interesting conclusions in the area of stability and frequency response.

The frequency response of a conventional PLL is determined by the loop filter, which in general has low pass characteristics. At best, the design of the loop filter is a compromise between obtaining a wide enough bandwidth that the loop will follow the modulation and a narrow enough filter that the carrier itself is somewhat rejected. The *order* of a conventional loop is determined by the number of poles in its response, particularly in its filter. A first-order loop (ie, one pole) is obtained with no filter at all. (But the loop response contains a pole because of the integral relationship between the frequency and the phase of the VCO.) If a simple low-pass loop filter is used, resulting in a second-order loop, and if the filter cutoff is set at a low frequency, the loop response will be narrow band. The loop will be unable to track a rapidly changing input. Further, as loop gain is increased, the response is underdamped, and the transient response is poor.

Lead compensation is normally used to solve this problem. A filter with a lag-lead characteristic results in a second-order loop with improved transient response and stability. However, the carrier rejection is degraded, and additional filtering may be required outside the loop.

Our nonlinear system is slightly different. With the zero-order estimator used in the phase detector, the loop response appears to be intermediate between first and second-order. At low modulating frequencies, the phase detector is effectively instantaneous, and since there is no filter, the loop response is apparently first-order. At high modulating frequencies, the time lag in the steps of the phase detector output makes the response approximate a second-order loop. As the loop gain is increased, there is a decrease in the damping. In fact, to some extent the loop gain can be used to tailor the frequency response. A discussion of feedback systems with zero-order hold is given by Elgerd (1967) and by Kuo (1967).

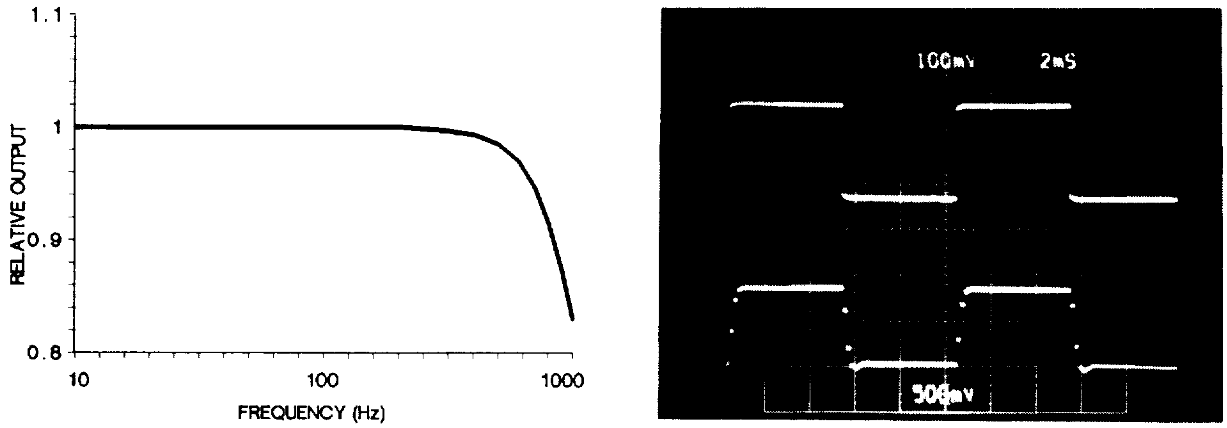
With our Type III system, analog filtering need not be used, although the possibility is not ruled out. Another possibility is the use of a stepwise signal for lead compensation derived in the same way as the signal for the difference ramp generator in the first order estimator. This could be applied in a summing circuit to tailor the loop frequency response. It is likely that the loop response would resemble that of an analog second-order loop with lead compensation. We have made no attempt to investigate such a loop in our prototypes, so far. We note, however, that there is a measure of lead compensation inherent in the first-order estimator.

2.7. Performance evaluation

While the waveforms and spectra presented above are indicative of improved performance, the usual way of assessing any improvement is in terms of the frequency response (or the step response) and the dynamic range of the link. In this section, we examine those aspects of the performance of a link using the new phase-locked loop. In these examples, the carrier frequency is about 5 kHz.

2.7.1. Frequency and step response

The frequency response and the step response of a link using the zero-order loop are shown in Figure 2-13.



(a) Frequency response (b) Step response

Figure 2-13. Frequency and step response, zero-order loop

The high frequency response can be adjusted between a peak and a roll-off by means of the loop gain. The step response, shown in Figure 2-13b, shows both overshoot and undershoot, perhaps evidence of nonlinear behavior. The low frequency response is excellent, extending (as with conventional FM) down to dc. The high frequency response was found to be slightly amplitude dependent. It could be peaked, in the region of 1 kHz, by changing the loop gain.

2.7.2. Dynamic range

The dynamic range is shown in Figure 2-14, which was plotted at a frequency of 80 Hz, in the linear range of operation, and close to the 60-Hz that is of interest in most power system measurements. The transfer function of the zero-order loop with a simple two-pole low pass filter is within <10% from about 0.5 V down to a level of 200mV as shown.

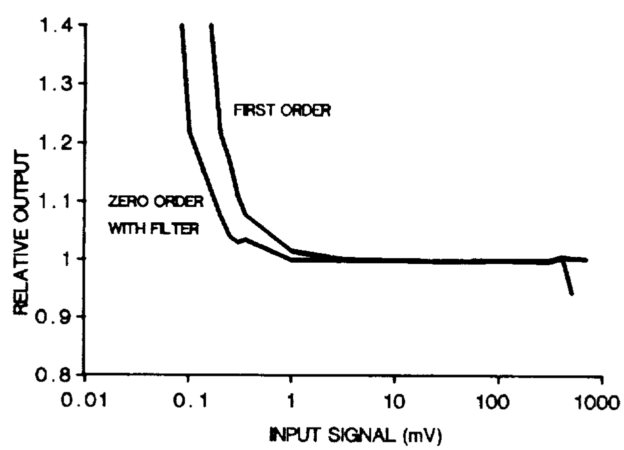


Figure 2-14. Dynamic range

The small "glitch" in the characteristic near 300 μV is almost certainly an artifact of our measurements. The system has a dynamic range in the order of 70 dB, enough to make it useful in a number of measurement applications, but also enough to make characterization difficult. During the course of these measurements, it was necessary at various times to add attenuators (at the input) and an amplifier (at the output) and to change range on our measuring voltmeter. The sudden apparent jump in the response corresponds to one of these changes.

2.7.3. Lock range

The input-frequency/output-voltage relationship of a phase-locked loop is typically a sawtooth. As the frequency is increased slowly from an unlocked low frequency, the loop provides no output. The average output of the phase detector, filtered in the normal PLL, is zero. At some point, the loop will lock, and the output suddenly jumps to a large value. This is the value needed to pull the VCO to the input frequency. As the frequency continues to increase, the output voltage will track the frequency (providing a way to measure the linearity of the PLL) until, at some point, the loop loses lock, and the output voltage returns to zero. If the input frequency is now slowly decreased, the loop will reacquire lock, usually at a lower frequency than the frequency at which it lost lock. Lock will be maintained, usually to a lower frequency than the frequency at which lock was first acquired.

The foregoing description applies to a loop with a Type I phase detector. The slope of the characteristic is the reciprocal of the VCO gain. The Type II detector will give the loop similar performance, but will often have a wider lock range, since the Type II phase detector is really a phase/frequency detector, and can apply a non-zero voltage to the VCO even when the loop is not locked. The test was performed on the Type III loop, with the results shown in Figure 2-15.

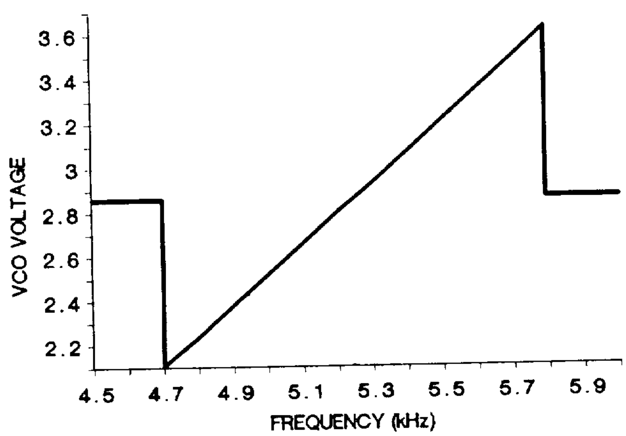


Figure 2-15. Lock range

The range for lock and acquisition are indistinguishable. The linearity is excellent, in fact the nonlinearity is observed to be <0.3%, the uncertainty of our measurements. The lock range is large enough, in terms of the input frequency, to make the device useful as a tachometer.

2.8. Concluding Remarks

Detectors for FM are not new. Indeed, there are a large number of ways that an FM signal can be modulated and demodulated, the choice being dependent on the application. For use in commercial VHF radio, the "discriminator" was once the detector method of choice. Recently, the phase-locked loop has been used in this application. Since the required output is band-limited at less than 20 kHz, it is not difficult to remove the PLL phase detector products, which are all above the IF of 10.7 MHz. As with conventional discriminators, a simple passive single-pole filter will perform well.

The constraint in our application is primarily at the modulator. The uncertainties in the power margin of the optical data link lead to the choice of frequency modulation, since the output is independent of the input signal level. The limited power available in our optically-powered sensors then leads inevitably to systems with very low center frequencies. For a fixed (optical) pulse width, the power consumption of the optical driver stage is proportional to frequency.

As the carrier frequency is lowered, to conserve power, the filtering problem becomes more difficult to solve. Desired output signals can have frequencies that are an appreciable fraction of the carrier, and of undesired phase detector products.

The demodulator described above solves the filtering problem by greatly reducing the unwanted phase detector outputs. This simplifies the design of the demodulator end of the link, and results in excellent overall system performance.

While first-order designs have been evaluated, and higher orders are possible, in future optical links that use FM, our group is likely to use the Type III zero-order PLL design described above, perhaps combined with a two-pole filter. This seems to have the best combination of performance and simplicity of the FM systems considered.

3. TEMPERATURE EFFECTS

During evaluation of the prototype optically powered link, variations in the signal amplitude as a function of temperature were noted. For the data link to be of most use, these temperature effects must be understood, and compensated.

During the development of our earlier optically powered link, it was noted that the temperature-dependence of the frequency was a function of the value of the resistor used to determine the frequency. This resistor also affected the current in the VCO. It happened that when the resistor was chosen for a low value of current, to meet the limitation of low power, the oscillator was quite temperature dependent. Representative results are shown in Figure 3-1.

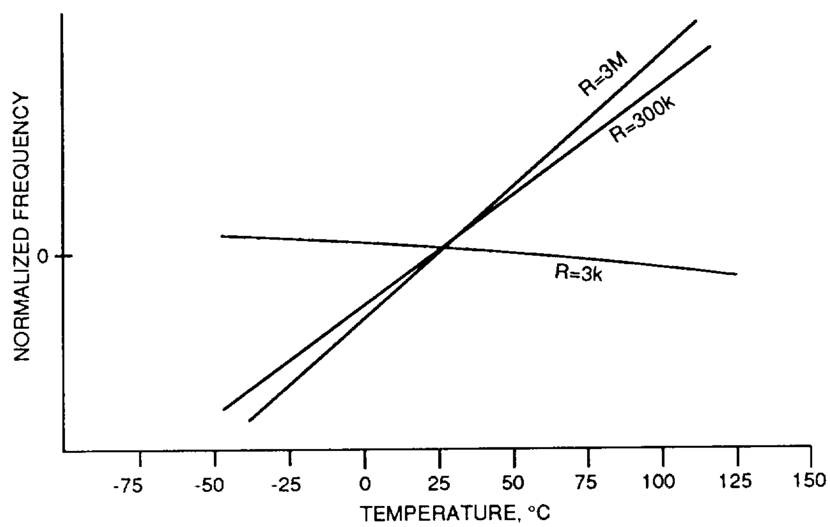


Figure 3-1. Resistor-dependence of temperature effects, CD4046

Clearly, the marked temperature dependence evident at most values of resistance and temperature should be compensated. There are two constraints on the temperature compensation method used. It must not affect the sensitivity of the VCO, and it must have very low current consumption. The method we used takes advantage of a frequency offset capability of the CD4046.

To understand this properly, one must understand how the VCO in the CD4046 operates. The control voltage is applied to a FET, in the source-follower mode. This voltage therefore appears across the timing resistor R_1, as shown in Figure 3-2, fixing its current. This current is arranged to be part of a current mirror circuit, controlling the charging current into the timing capacitor.

The current mirror can also be programmed directly. Any current into (or out of) pin 12 of the device adds to (or subtracts from) the current through the timing resistor, and changes the charging current by means of the current mirror. Originally designed to have an external resistor R_2 to control the programmed offset current, the circuit works well with an external current source connected instead.

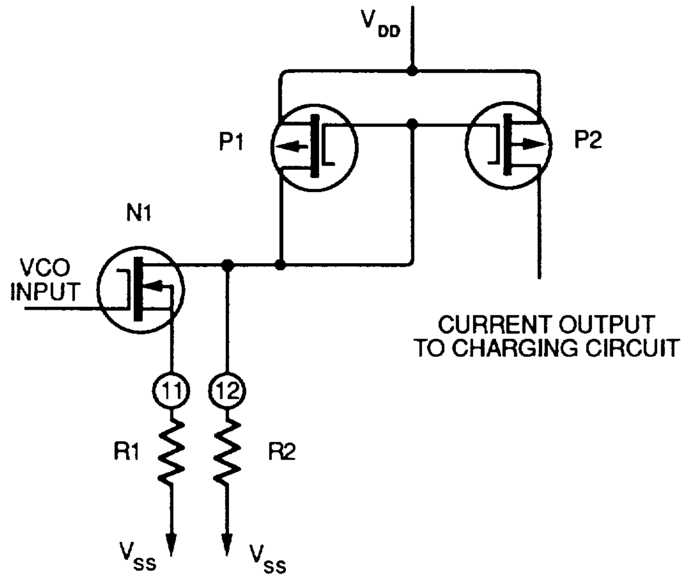


Figure 3-2. CD4046 internal current mirror timing circuit

To achieve the desired temperature compensation, a temperature-dependent current was used to program the internal current mirror. In Figure 3-3, transistor Q_2 and diodes D_1 to D_3 form a current sink whose current is set by the value of R and the forward volt-drops of the diodes and the base-emitter junction of the transistor. These volt-drops are temperature dependent (about -2.1 mV/°C each) so that the current decreases with temperature. This is in the right direction to compensate the CD4046 oscillator, but there is a large component of the current that is not temperature dependent, because of the ≈0.65-V drop of a forward biased junction. This component of current is removed by transistor Q_3, which is operated as a current source. Diode D_4 compensates (approximately) for the temperature-induced variations in the V_{BE} of Q_3, so that the programmed source current is almost independent of temperature.

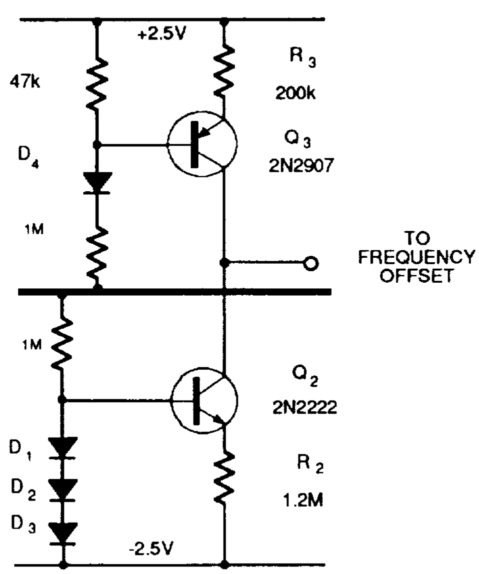


Figure 3-3. Temperature compensation circuit for VCO

The compensation circuit is connected to the frequency offset terminal (pin 12) of the VCO. Any difference between the temperature-dependent source and the constant sink currents must be made up by a current source inside the CD4046 circuit. It adds algebraically to the timing resistor current. The use of current compensation (rather than resistive) means that the offset is independent of the voltage on the VCO input, ie, it does not change the sensitivity of the VCO.

By means of this configuration, the VCO free-running frequency can be made independent of temperature over quite a wide range. Of course, adjustment of the compensation current affects the frequency, but since there are now three parameters that can be controlled (resistance, capacitance and compensation current), it is possible to choose a resistor for low-power circuit operation, a compensation current for zero temperature dependence, and a capacitance to set the frequency.

It was therefore somewhat perplexing to find during temperature tests of a hybrid integrated circuit version of the link, used with an integrator, that the amplitude at the receiver was not constant. The effect was attributed to temperature dependent changes in the value of the integrator capacitor. Figure 3-4 shows the effect. While a gain change of 0.5 dB due to this mechanism would ordinarily be unimportant, in a measurement system this represents a 5% error.

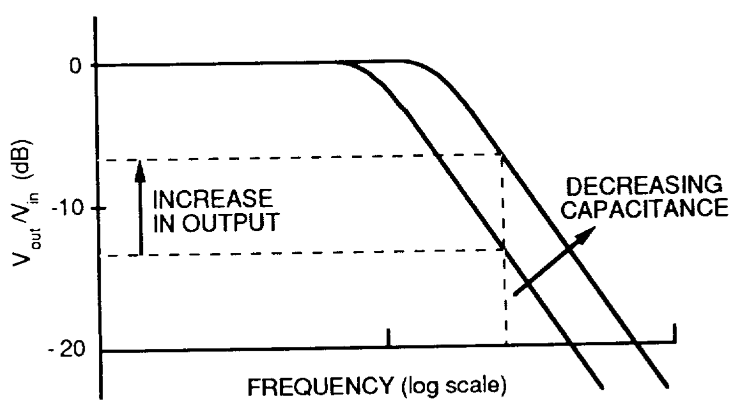


Figure 3-4. Integrator characteristics

Several measurements require the use of an integrator. For example, the measurement of magnetic field by means of the voltage induced in an open circuit coil, the measurement of electric field by means of the current between two short circuited electrodes, and the measurement of current by means of a linear coupler all require that the measured signal be integrated. We think that the linear coupler approach shows considerable promise in distribution automation, so the integrator problem was investigated further.

A constraint is again applied by the use of optical power, and a further constraint may be encountered in field measurement work. The low power problem may rule out the addition of complex temperature compensation circuits, and the application to field measurements is likely to be driven by size goals, again ruling out additional circuitry. Therefore, means of compensating the transmitter by changes in the receiver were sought.

There are two parts to the solution. First, the transmitter is deliberately decompensated so as to make the free-running frequency act as a surrogate for the temperature at the transmitter. This necessarily implies that the link cannot be used at dc. In most power system applications this is not a serious limitation. Second, the receiver gain is made to depend on this indication of remote temperature so as to correct for the observed gain changes. It is assumed that the temperature dependence of the integrator capacitor is a simple function, that can be adequately corrected for by simple compensation. Our experimental results confirm this; it seems that straightforward linear temperature correction is adequate.

It is easy to decompensate the transmitter, so as to make the free-running frequency of the VCO any desired function of temperature. Simply by adjusting the compensation current (into pin 12), the temperature dependence can be made positive, zero or negative, over quite a wide range. The diode volt-drops (D_1 to D_3 in Figure 3-3) vary uniformly with temperature. The compensation current is therefore a linear function of temperature. Although the timing resistor value is chosen for low current consumption, and--uncompensated--results in a nonlinear temperature dependence, the overall effect is that the frequency is almost a linear function of temperature, see Figure 3-5.

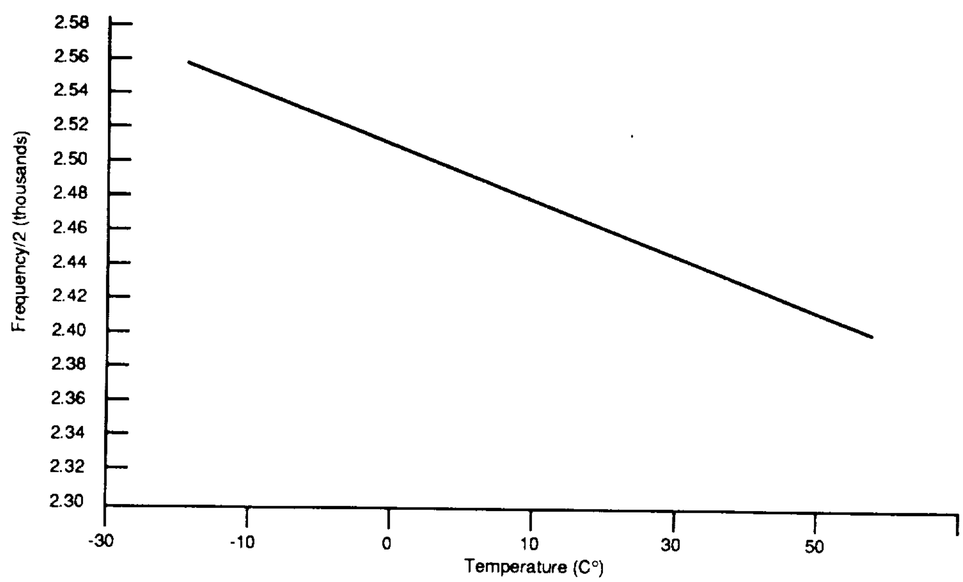


Figure 3-5. Decompensated VCO characteristics

At this point in the development, we have a transmitter whose gain is a function of temperature, and whose free-running frequency is a (controllable) function of temperature. It is possible that the VCO transfer characteristics, which depend on free-running frequency, are now also temperature dependent. This effect is due to the change of frequency, and not the choice of timing components. The timing components chosen have little or no effect on the slope (or the capture range) for a constant frequency. This can be seen in Figure 3-6, which compares a VCO performance for two values of timing capacitance, C_t with the timing resistor adjusted for constant frequency.

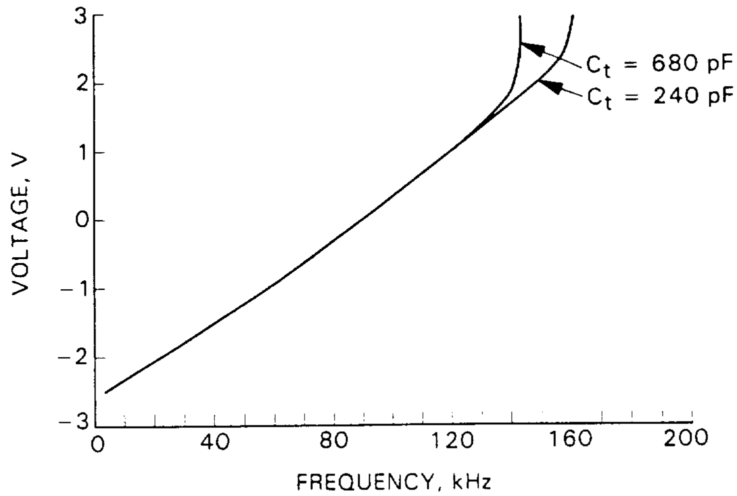


Figure 3-6. Effect of timing component values on conversion characteristics

As the center frequency, f_c, was increased by changing either the timing resistor or capacitor, the slope of the curve decreased. This change appears to be a linear relationship. Figure 3-7 demonstrates this.

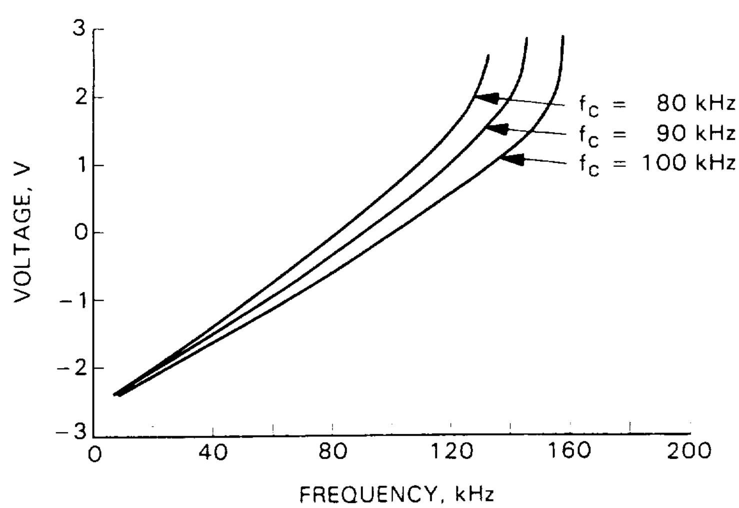


Figure 3-7. Change in slope with center frequency

Thus, it is anticipated that transmitter transfer characteristic temperature dependence is an effect that can be corrected, since it may not be distinguishable from the gain changes in the integrator.

In the receiver, then, the link signal and the transmitter center frequency must be separated, and the gain of the signal measurement system made to depend on the frequency. Figure 3-8 is a block diagram of the arrangement used to accomplish this.

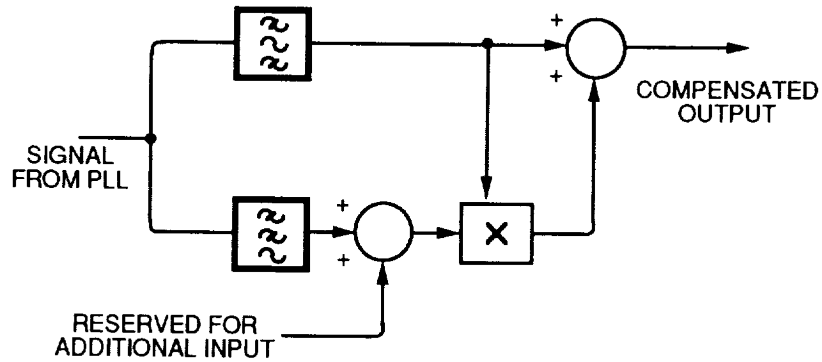


Figure 3-8. Block diagram of gain compensation system

The signal from the detector is shown as being applied to two separate filters. One is the normal signal processing filter, if used. This is shown as a band-pass filter: the signal is assumed to be ac coupled. The other is a low-pass filter with a very low cutoff frequency (possibly minutes) that essentially recovers the temperature effects at the transmitter, and removes the modulation. This signal, the dc level of the phase-locked loop demodulator, represents the free-running frequency of the transmitter (assuming that the *average* value of the modulation is zero), and hence reflects the temperature, and hence the gain. By multiplying this information by the demodulated signal, a compensation signal can be derived. An analog multiplier is a convenient way to do this. This compensation is added to the original signal, so that the system output is corrected for transmitter temperature effects.

In the preceding discussion, the receiver demodulator was assumed not to be temperature sensitive. In practice, it is not necessary to achieve this perfectly. The temperature compensation system shown in Figure 3-8 can be used to correct for temperature effects in the receiver, too. In this case, a temperature measurement circuit, possibly using diodes to sense the local temperature, provides an additional input to the multiplier.

Uncompensated, the gain of a typical system might vary 10% between 0°C and 40°C. If the compensation is defined as zero at 20°C, it must then reach a maximum of 5% at the two extremes of temperature. It is not difficult to adjust the compensation so that the overall gain is stable within 1% over the range. This amounts to a drift of 0.025%/°C, or 250ppm/°C. With careful adjustment, and patient use of the temperature-controlled environment, this figure can be improved.

4. A PERIOD-MODULATING SYSTEM

4.1. Background

A second development is concerned with telemetry of analog information. In telemetry over optical fibers, it happens that representing an analog quantity by means of an analog light level is usually unsatisfactory, and alternative means are used. A commonly used alternative is to employ frequency modulation (FM), since the detection process is then essentially insensitive to the amplitude of the received signal.

Detection (or demodulation) of an FM signal can be accomplished in a number of ways. At high carrier frequencies, a discriminator may be used. Over a wide range of frequencies, a phase-locked loop may be used. At low carrier frequencies, better rejection of the carrier is obtained by a system that combines the techniques of state estimation and phase locking to cause a local oscillator to mimic the frequency modulated oscillator. (This last approach was described in Section 2 of this report.)

The present development retains the state estimation to improve the carrier rejection, but it uses a different modulation technique, so that linearization by means of a phase-locked loop is not needed.

4.2. Description of telemetry link

There are several ways to implement a voltage controlled oscillator (VCO). At high frequencies, it is usual to control the value of a circuit parameter, such as a capacitance or a mutual inductance, by means of the applied control signal. At low frequencies, it is more common to employ a relaxation oscillator, and to have the control signal control the charging rate of a capacitor in the circuit. In such a system, the capacitor will charge until it reaches a specific target voltage, at which point it is discharged, and the process repeats. If the charging rate is increased, the target is reached sooner, and the frequency is increased. Systems like this can implement very linear frequency modulation.

In the usual phase-locked loop application, the modulation is performed by a voltage controlled oscillator (VCO): the output *frequency* is determined by the input voltage. The period between pulses, the *only parameter that can be directly measured using only two pulses*, is instead the *reciprocal* of the frequency. This seemingly intractable mathematical identity has not been resolved in any way that is widely accepted as satisfactory.

The prototype of the new system uses a relaxation oscillator, but the charging rate is held constant and the target voltage is controlled. Now if the target voltage is increased, the capacitor takes longer to reach the value, and the frequency is decreased. By analogy with FM, we may call this period modulation. The effect is the reciprocal of FM. Figure 4-1 contrasts the two modulation methods.

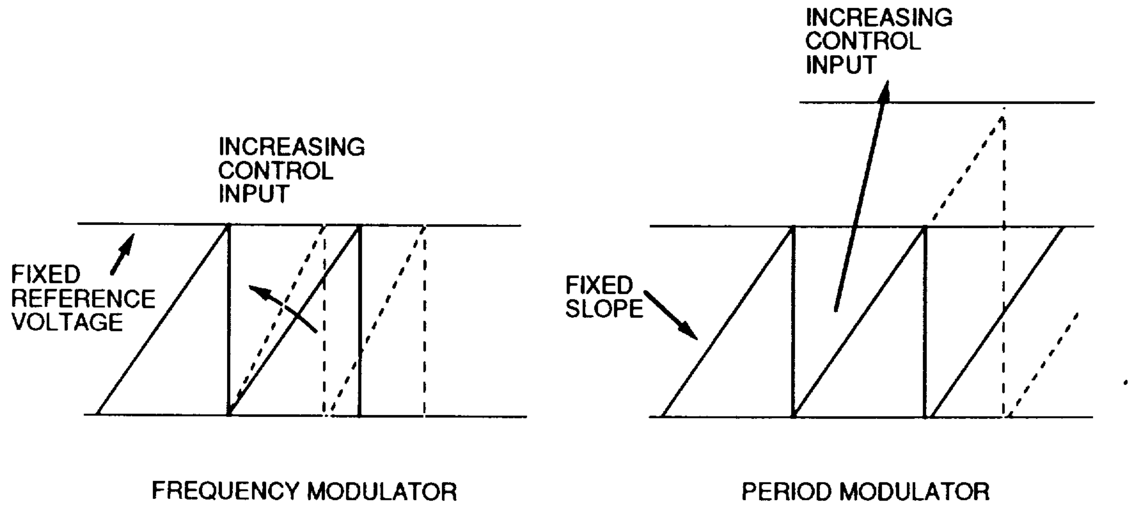


Figure 4-1. Comparison of frequency modulation and "period" modulation

The modulation process can be replicated in the receiver circuit, and the modulating signal recovered efficiently. If the state estimation method of demodulation is used, the output of the sample-hold circuit is an estimate of the modulating signal, without the need to enclose the estimator inside a phase-locked loop. Zero- or first-order estimators have been built in our prototypes; higher orders may be possible.

What one would ideally like is a modulating/demodulating system in which it is easy in the demodulator to reproduce what is going on in the modulator. An important result of this new system is that if the period is modulated instead of the frequency, recovery of a high quality estimate of the modulation is greatly simplified.

The prototype modulator is shown in the block diagram Figure 4-2. The system functions as follows. The ramp generator produces a linear ramp with a constant slope. The output of the generator is applied to one input of a comparator. The control voltage, instead of controlling the ramp slope as in FM, is applied to the other input of the comparator. When the ramp voltage reaches the value of the control voltage, the comparator changes state. This causes the ramp generator to be reset, which reverses the state of the comparator. The process repeats, causing a train of period-modulated pulses to be generated.

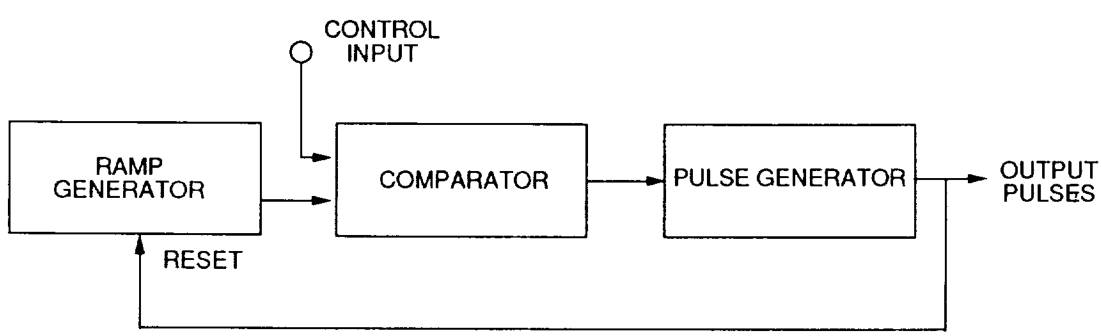


Figure 4-2. Prototype period modulator

The demodulation process is fundamentally a replica of the modulation process. In both the modulator and the demodulator, the pulses correspond to the starting and stopping of the ramp generator. The result is that the ramp generator in the demodulator mimics exactly the ramp generator in the modulator. (This is a clue to the excellent linearity of the system.) As a practical matter, there is a very small delay because of the finite width of the reset pulses.

The information conveyed by the arrival of a pulse from the modulator is precisely that at this instant, the signal voltage was equal to the ramp voltage. Since we know when the previous pulse arrived, and how fast the modulator's ramp generator charges, we know the value of the signal voltage at the instant of each pulse. In fact, since we have a local ramp generator that mimics the one in the modulator, we have a replica of this voltage in the demodulator.

In terms of signal recovery, we can estimate the signal between pulses. The zero-order estimate is that the signal was constant in the interval. A zero-order hold circuit operating on a sample of the ramp voltage immediately before the arrival of each pulse implements this. The first-order estimate is that the rate of change of the signal was constant in the interval. This is implemented[3] by a first-order hold circuit. In our prototype link we used the sample-hold system described earlier to create the state estimator circuit. The arrangement is shown in block diagram form below.

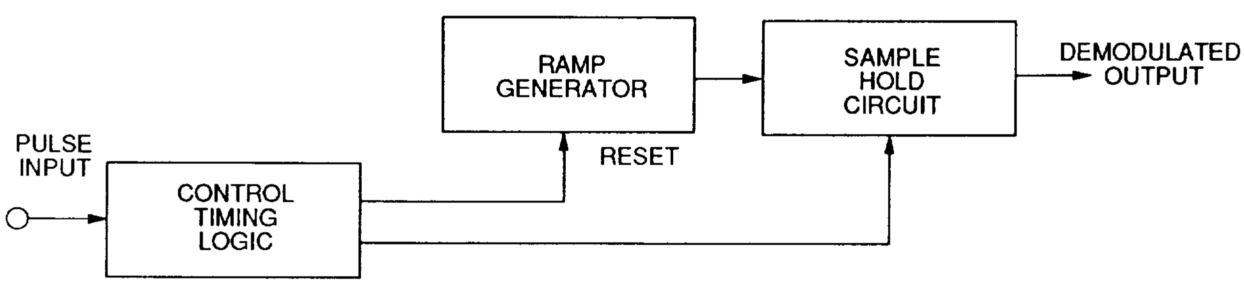


Figure 4-3. Prototype period demodulator

As a practical matter, it is important to ensure that the signal voltage applied to the comparator in the modulator does not reach zero, or more specifically, the discharge value of the ramp. This would force the oscillator to infinite frequency, a practical impossibility. In our implementation, there would be some nonlinearity before this, because of the finite width of the reset pulses, so there is further reason to ensure the condition is not reached. It is a simple matter to offset the control voltage before it is applied to the comparator, and to remove the offset in the demodulator. A bandgap reference diode provides a convenient and temperature stable means of offsetting both modulator and demodulator. The results shown below were obtained with a system of this kind.

[3] Approximately. As with the Type III phase-locked loop described earlier, the ramp generating circuit that generates the signal that converts a zero-order estimate into a first order one is accurate only if the period between pulses is constant. The fact that it is not gives rise to some of the observed nonlinearities of the system.

An alternative implementation in which the ramp generator operates between the supply rails rather than ground and one rail has not been tested, but it is clear that it would obviate the need to offset the signal. Since the ramp and the signal could each be almost as large as the total supply voltage, it is likely that there would be an improvement in the dynamic range.

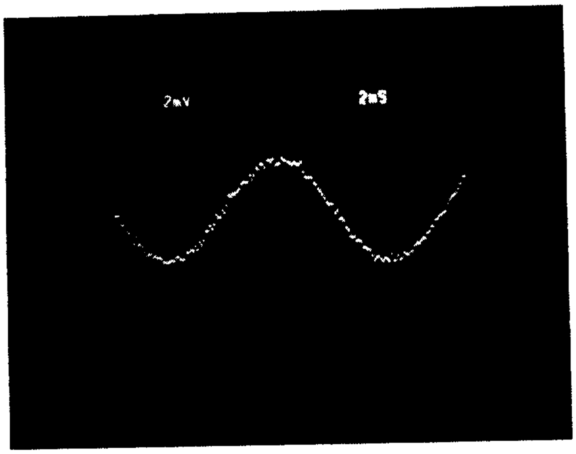
A further small refinement overcomes a possible problem with the modulator. If the reset pulse for the ramp generator is triggered by a change of state in the comparator, a single missed reset can cause the system to "lock up." Once the comparator has gone into the state that indicates the ramp voltage exceeds the input, it will not change state again. If, for some reason, the ramp failed to reset, no further output pulses would be generated. This condition can be avoided by having the reset pulse generated by the state of the comparator, rather than its transition, or by means of a "watchdog" circuit that generates an additional pulse if too much time has elapsed between reset pulses.

4.3. Period modulator—measured results

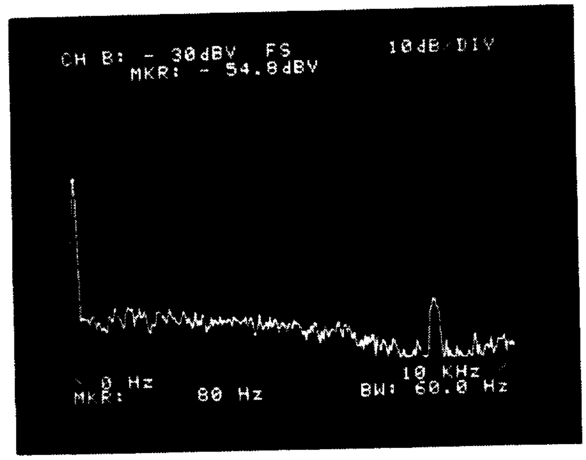
Because the period demodulator uses the state estimator approach, it is to be expected that its performance resembles that of the phase-locked loop described in Section 2. This is broadly true, though in some respects it outperforms the PLL system.

Figure 4-4 shows the output signals and the spectra for input signals of 2, 20 and 200 mV at 80 Hz. (These results may be compared with Figures 2-7, 2-8 and 2-9.) In each case the spectrum shows the signal to be 40 dB or more above the fundamental of the carrier. At 2 mV (Figure 4-4 (a)) the discrete nature of the output is evident. Some uncertainty in the levels of the samples can be seen. This may be due to amplitude noise in the modulator comparator or demodulator sample-hold, or timing jitter.

At the higher signal levels (Figure 4-4 (c)-(f)) there seems to be less evidence of harmonic distortion than in the case of the phase-locked loop. This is particularly noticeable in the spectrum of the signal corresponding to a 200-mV input, Figures 2-8 (f) and 4-4 (f). In the case of the PLL, some output at harmonics of the 80-Hz carrier is clearly evident, although to be sure their level is in the order of 60 dB below the fundamental. Harmonic distortion products appear to be totally absent from the spectrum of the period modulation system.

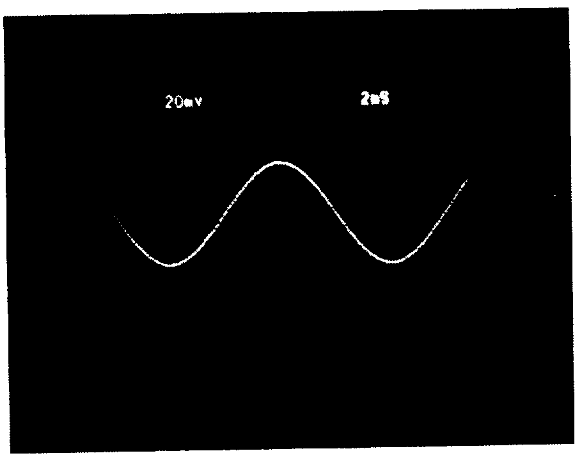


(a) Output signal

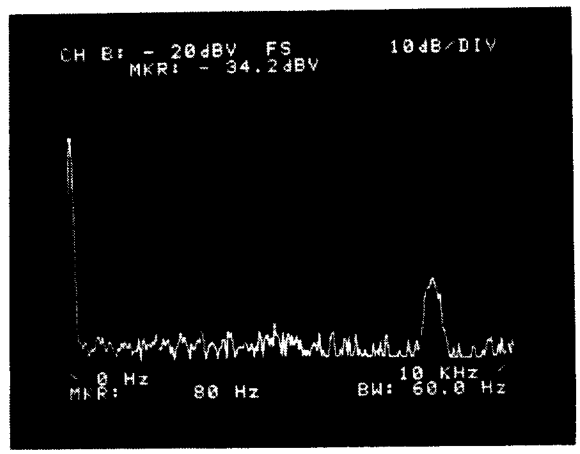


(b) Output signal spectrum

Figure 4-4 (a) and (b). Period modulator/demodulator performance, V_{in} = 2 mV

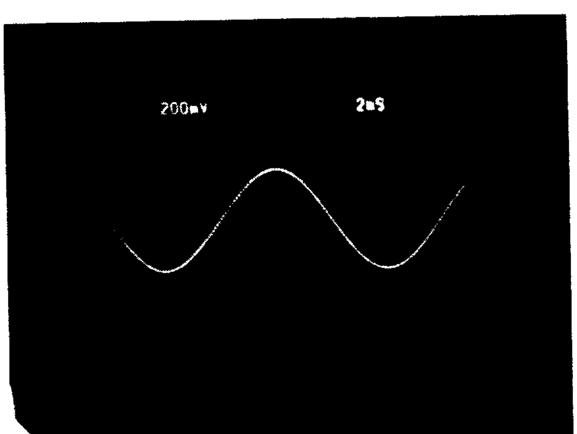


(c) Output signal

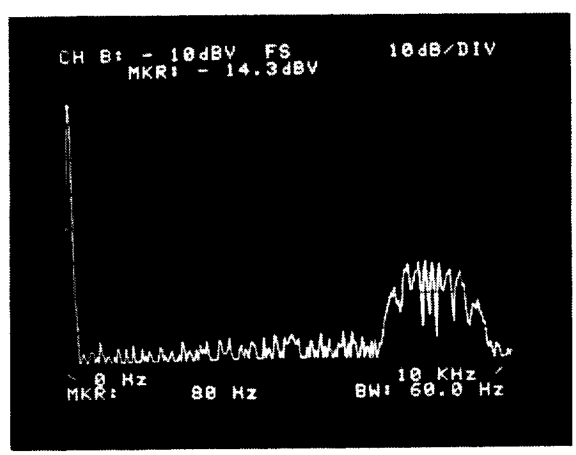


(d) Output signal spectrum

Figure 4-4 (c) and (d). Period modulator/demodulator performance, V_{in} = 20 mV



(e) Loop output signal



(f) Output signal spectrum

Figure 4-4 (e) and (f). Period modulator/demodulator performance, V_{in} = 200 mV

The dynamic range of the system is shown in Figure 4-5 for a signal frequency of 80 Hz.

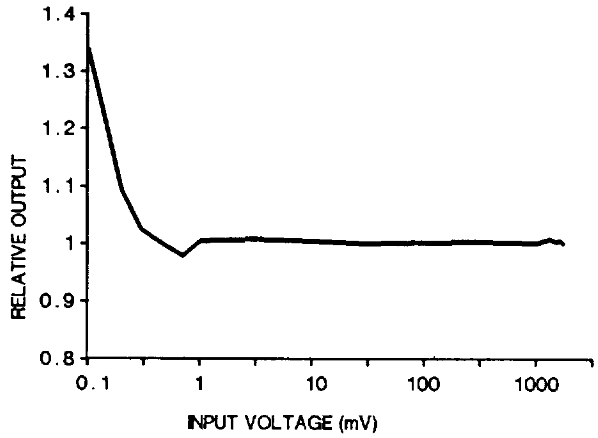


Figure 4-5. Period modulator/demodulator dynamic range

As expected, the dynamic range is comparable with the PLL system, and in the order of 70 dB. The discontinuity in the output near an input of 1 mV is due to the insertion at this signal level of an amplifier (external to the period modulator/demodulator system) to facilitate the measurements.

The frequency response is excellent, as shown in Figure 4-6.

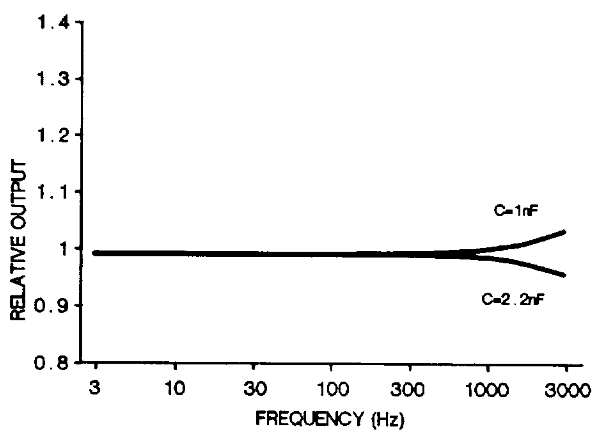
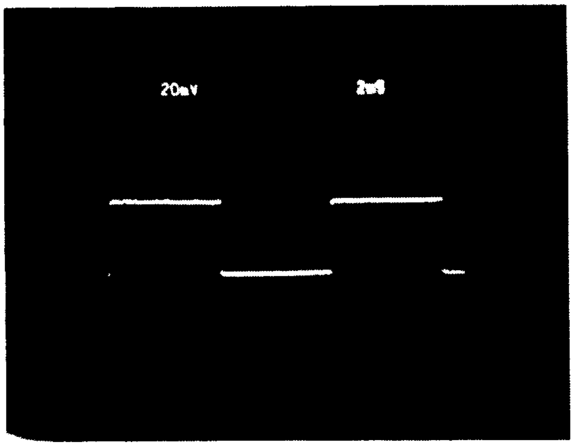


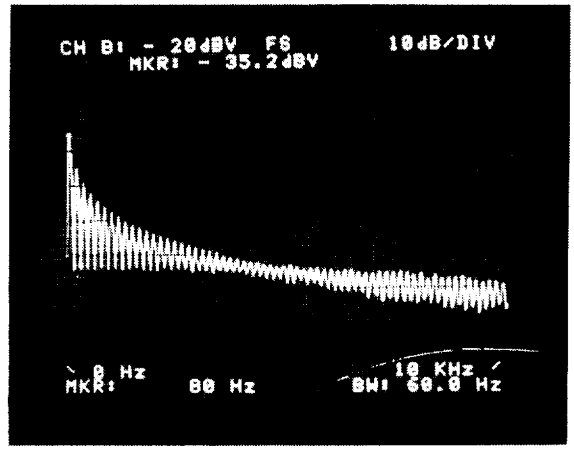
Figure 4-6. Period modulator/demodulator frequency response

Above about 1 kHz there is a slight increase in the output, or a decrease, depending on the size of the capacitor in the receiver sample-hold circuit. This distortion is quite unimportant, as there is in any case a frequency aliasing problem at such high ratios of frequency to sample rate. The frequency response of the period modulated link was not particularly amplitude dependent.

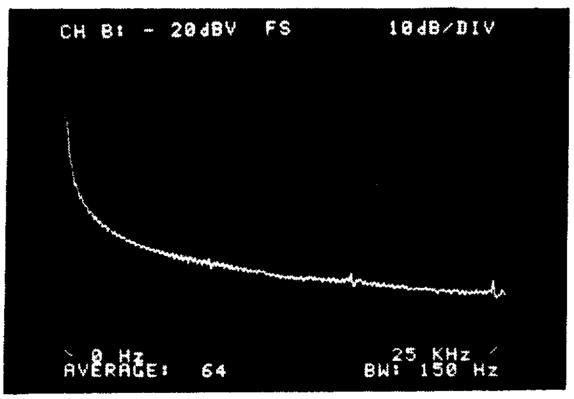
ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH



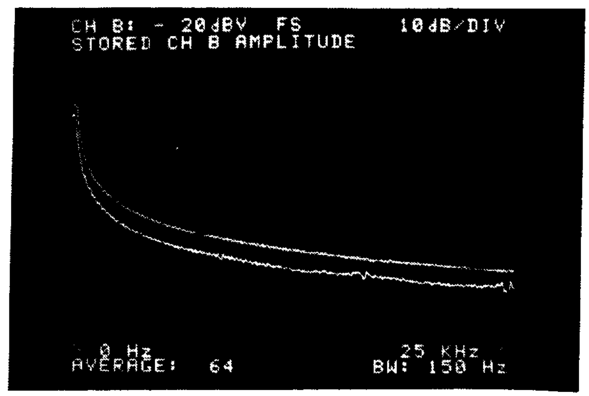
(a) Output signal



(b) Output signal spectrum



(c) Output spectrum to 25 kHz



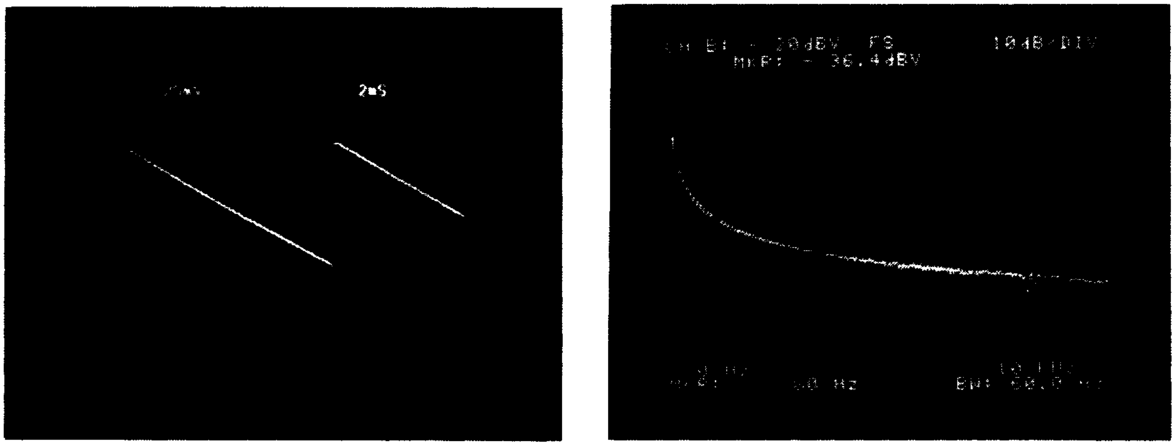
(d) Input signal spectrum superposed

Figure 4-7. Square wave response

The spectrum plots are a little misleading, in that the harmonics seem to be asymptotic to some value. Actually, this effect is an artifact of the spectrum analyzer, which has a linear abscissa. The amplitude of the 10-kHz harmonic is 20 dB down on the 1-kHz figure, and the 25-kHz response is 20 dB down on the 2.5-kHz output, indicating the expected 1/f decrease in harmonic output associated with a square wave. This is verified in Figure 4-7 (d), in which the spectrum of the input signal is included. The two spectra are very similar.

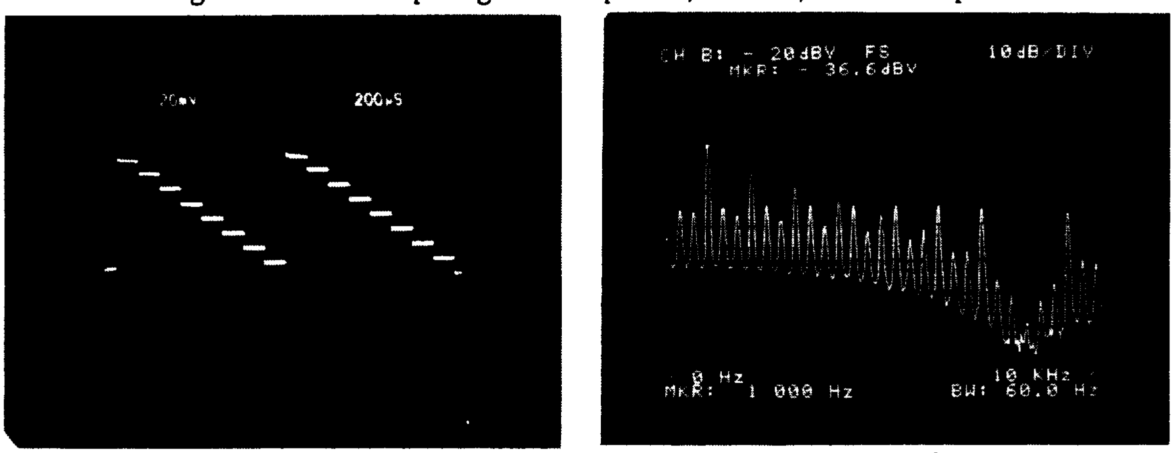
In some applications of the system as a link, there might be interest in the response to ramp signals. For example, the fields associated with video display terminals (VDTs) may be ramplike, because of the need to scan the screen. Figures 4-8 and 4-9 show the link response to ramps.

The ramp response is clearly as good as the square-wave response. At an input frequency of 1 kHz, there are only 8 or 9 samples per cycle, but within this limitation, the output is a faithful copy of the input. If the system were to be used to measure VDT fields, with a repetition rate of about 15 kHz, a carrier frequency in the order of 100 kHz should suffice for a reasonably accurate measurement of the harmonic energy.



(a) Output signal (b) Output signal spectrum

Figure 4-8. Ramp signal response, 80 Hz, 20 mV input



(a) Output signal (b) Output signal spectrum

Figure 4-9. Ramp signal response, 1 kHz, 20 mV input

4.4. Concluding Remarks

The period modulator/demodulator system described above enables a telemetry link of low carrier frequency to reproduce low-frequency signals accurately. Harmonic-rich signals, such as might be generated in phase-controlled circuits, and ramp signals, such as those caused by screen scanning circuits, are also transmitted accurately. The system also exhibits good dynamic range without the use of filtering.

A zero-order design has been evaluated, and while higher orders are possible, it may be inferred from the results presented in Section 2 that the extra complexity is not usually warranted by the improvement in performance. In future optical links our group is likely to use the zero-order period modulation system described above in preference even to an FM system using the Type III phase detector described in Section 2.

5. REFERENCES

Adolfsson, M., Einvall, C.-H., Lindbergh, P., Samuelsson, J., Ahlgren, L. and Edlund, H. (1989, April). EHV Series Capacitor Banks. A new Approach to Platform to Ground Signalling, Relay Protection and Supervision. IEEE Transactions on Power Delivery, 4(2), pp. 1369-1378.

DeLoach, B.C. and Gordon, E.I. (1981, October). Optically powered implant in the human body. Western Electric Technical Digest (No. 64), pp. 17-18.

DeLoach, B.C., Miller, R.C. and Kaufman, S. (1978, November). Sound alerter powered over an optical fiber. Bell System Technical Journal (No. 57), pp. 3309-16.

Elgerd, O.I. (1967). Control Systems Theory, New York: McGraw-Hill.

Gardner, F.M. (1966). Phaselock Techniques, New York: Wiley & Sons.

Hall, P. (1986). An optically powered sensor network. IEE Colloquium on Distributed Optical Fiber Sensors. (Digest No. 74).

Holte, K.C. (1989, July). Technology Requirements for a Competitive Electric Utility Market in the 21st Century. IEEE Power Engineering Review. 9(7), pp. 18-22.

Hopper, A., Temple, S. and Williamson, R. (1986). Local Area Network Design Wokingham, England: Addison-Wesley.

Kirkham, H. and Johnston, A.R. (1988). Electric and Magnetic Field Meters developed for the U.S. Department of Energy (JPL Publication 88-1). Pasadena, CA: Jet Propulsion Laboratory.

Kirkham, H., Johnston, A.R. and Friend, H.A. (1989). Distribution Automation Applications of Fiber Optics (JPL Publication 89-10). Pasadena, CA: Jet Propulsion Laboratory.

Kirkham, H., Johnston, A.R., Lutes, G., Daud, T. and Hyland, S. (1984). Power System Applications of Fiber Optics (JPL Publication 84-28). Pasadena, CA: Jet Propulsion Laboratory.

Kuo, B.C. (1967). Automatic Control Systems (Second Edition). Englewood Cliffs, NJ: Prentice-Hall.

McGlade, S.M. and Jones, G.R. (1984). An optically powered vibrating quartz force sensor. G.E.C. Journal of Research. (U.K.) 2(2) pp. 135-8.

Ohte, A., Akiyama, K. and Ohno, I. (1984). Optically-powered transducer with optical fiber data link. SPIE. 478(Fiber Optic and Laser Sensors II) p. 3338.

Schweizer, P., Neveux, L. and Ostrowsky, D.B. (1987). Optical fiber powered pressure sensor. SPIE. 798(Fiber Optic and Laser Sensors II) pp. 82-5.

Zimmermann, H. (1980, April) OSI Reference Model - the ISO Model of Architecture for Open Systems Interconnection. IEEE Transactions on Communications COM-28(4) pp. 425-432.

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No. JPL Pub. 90-2	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Developments in Fiber Optics for Distribution Automation		5. Report Date December 1, 1991
		6. Performing Organization Code
7. Author(s) H. Kirkham, H. Friend, S. Jackson, and A. Johnston		8. Performing Organization Report No.
9. Performing Organization Name and Address JET PROPULSION LABORATORY California Institute of Technology 4800 Oak Grove Drive Pasadena, California 91109		10. Work Unit No.
		11. Contract or Grant No. NAS7-918
		13. Type of Report and Period Covered JPL Publication
12. Sponsoring Agency Name and Address NATIONAL AERONAUTICS AND SPACE ADMINISTRATION Washington, D.C. 20546		14. Sponsoring Agency Code PL77812010203 (RE152)
15. Supplementary Notes		

16. Abstract

An optical fiber based communications system of unusual design is described. The system consists of a network of optical fibers overlaid on the distribution system. It is configured as a large number of interconnected rings, with some spurs. Protocols for access to and control of the network are described. Because of the way they function, the protocols are collectively called AbNET, in commemoration of the microbiologists' abbreviation Ab for antibody. Optical data links that could be optically powered are described. There are two versions: each has good frequency response and minimal filtering requirements. In one, a conventional FM pulse train is used at the transmitter, and a novel form of phase-locked loop is used as demodulator. In the other, the FM transmitter is replaced with a pulse generator arranged so that the period between pulses represents the modulating signal. Transmitter and receiver designs, including temperature compensation methods, are presented. Experimental results are given.

17. Key Words (Selected by Author(s))	18. Distribution Statement
Communications Electronics and Electrical Engineering Computer Programming and Software Optics	Unclassified -- Unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	64	